이학박사학위논문

Profiling and improving genome-wide specificity of CRISPR RNA-guided adenine base editors

2019년 8월

서울대학교 대학원

화학부 생화학 전공

김 다 은

Abstract

Profiling and improving genome-wide specificity of CRISPR RNA-guided adenine base editors

Da-eun Kim

Department of Chemistry

The Graduate School

Seoul National University

The CRISPR/Cas9 system is one of the most popular technologies in biotechnology field over the last decade. The CRISPR system is derived from the bacterial immune system and now functions as a programmable nuclease and a genome engineering tool. These genome engineering tools are used not only to knockout genes but also knockin with exogenous DNA by causing DNA double strand break and cellular repair system. So far, original CRISPR system has a limitation that it is mainly used for knockout because it is focused on non-homologous end joining (NHEJ) repair pathway rather than sophisticated homology directed repair (HDR).

However, about 58% of mutations known to cause human disease are point

mutations. Therefore, a technique for introducing a point mutation is required, and CRISPR RNA-guided base editors have recently been developed. CRISPR RNA-guided base editors are fused form of catalytically inactive Cas9 (Cas9 nickase) and cytidine/adenosine deaminases. Cytosine base editors (CBEs) can induce C:G to T:A and, adenine base editors (ABEs) can deaminase adenosine and convert A:T pair to G:C pair. These base editor systems have recently been reported in different types independently and are being applied to bacteria, human cells, animals, and plants, but there has been little research on specificity of base editors, especially adenine base editors.

In this thesis, I will first show that the substitution frequency of the ABE can vary with codon usage and the number of nuclear localization signals (NLSs). Then, I will introduce a method for verifying the off-target effects of ABEs by modifying Digenome-seq method, which is known as DSB recognition method. Off-target effects of ABEs are different from those of Cas9 nucleases or CBEs. I also demonstrate whether these off-target effects actually occur in human cells. Finally, I show that using modified sgRNAs, highly specific Sniper Cas9 nickase instead of WT Cas9 nickase, and ribonucleoprotein delivery methods in lieu of plasmids delivery are effective to increase specificity of ABEs.

Keywords: Genome engineering, CRISPR/Cas9, Adenine base editor, Genome-wide specificity, Modified single guide RNA, Sniper Cas9, Ribonucleoprotein delivery

Student Number: 2014-21245

Table of Contents

List of Figures

List of Tables

Ⅰ. Introduction

The human genome project (Cantor, 1990), which began in 1990, ended in 2003, faster than expected due to advances in sequencing technology and lower prices (Boeke et al., 2016). At last, it became possible to know the human 3.2 billion base pair sequence. Since then, as postgenomic era, research on functional genomics has been carried out to study the roles of each gene and mutations that cause disease, rather than merely reading the nucleotide sequence (Eisenberg et al., 2000). One of the most popular tools for studying this is the genome engineering tool, or programmable nuclease. Genome engineering tools are divided into three generations: the zinc-finger nucleases (ZFNs), the transcriptional activator-like effector nucleases (TALENs) and the clustered regularly interspaced short palindromic repeat (CRISPR)-associated proteins 9 (Cas9) RNA-guided engineered nucleases (Kim and Kim, 2014).

ZFNs are proteins consisting of zinc finger proteins (ZFPs) at the N-terminal and the catalytic domain of FokI restriction enzyme at the C-terminal (Kim et al., 1996). One finger motif of a ZFP recognizes 3-bp DNA sequences, usually 3-6 ZFPs linked to recognize 9-18 bp DNA. In addition, the catalytic domain of FokI acts as a dimer, so a pair of two ZFN units recognizes DNA that is 5-7 bp away and cleaves DNA double strand (Bitinaite et al., 1998).

The second tool, TALENs, is similar to ZFN in that it uses the catalytic domain of FokI as the nuclease domain. However, TALENs use the transcription

activator-like effectors (TALEs) derived from plant pathogenic *Xanthomonas spp.* bacterium as the DNA recognition domain (Boch et al., 2009; Moscou and Bogdanove, 2009). If the 12-13th amino acids of the repeat variable diresidues (RVDs) of TALEs are changed, they will recognize different DNA bases, and TALENs cleave the desired DNA target site as a pair (Boch et al., 2009; Moscou and Bogdanove, 2009).

The last programmable nucleases are CRISPR/Cas9. The CRISPR system derived from the bacterial or archaeal immune system is a system that plays a role in removing foreign nucleic acids such as DNA and RNA to prevent the re-infection of viruses (Barrangou et al., 2007; Makarova et al., 2006). The external nucleic acid is integrated into the repeat region on its bacterial own genome and is expressed in RNA form upon re-infection and acts with the related proteins to remove viral DNA or RNA. The most commonly used Type II CRISPR system consists of CRISPR RNA (crRNA) containing foreign DNA (protospacers), trans-activating crRNA (tracrRNA), and Cas9 protein that can cleave DNA. The most widely used Type II Cas9 is derived from *Streptococcus pyogenes* (SpCas9). It has a role of cutting double strands of DNA (dsDNA) if contains a 20-nt complementary sequence of crRNA and a 5'-NGG-3 PAM (protospacer adjacent motif) sequence at 3 ' (Hsu et al., 2014; Jinek et al., 2012). In contrast to ZFNs or TALENs, which need protein engineering for other target sites, CRISPR/Cas9 can target to a variety of genomes if only a short RNA sequence is changed. Since the report that CRISPR is used as a programmable nuclease *in vitro* in 2012 (Jinek et al., 2012) and that it is applicable in human cell (Cho et al., 2013; Cong et al., 2013;

Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2013; Mali et al., 2013), related studies have increased rapidly, and in *Science* journal, it is called CRISPR revolution (Barrangou, 2014).

These programmable nucleases serve to cleave the DNA located in the desired region of the genome. When DNA double strand breaks occur in the cell by the programmable nucleases, they are repaired by the repair mechanism of the cell itself. There are two major methods: nonhomologous end joining (NHEJ) and homologous recombination (HR). NHEJ is predominant in the cell as a result of the ligation of the cleaved DNA without a repair template. However, it is prone to error, like small insertions or deletions (indels) and frame shift, usually resulting in gene knockout. HR is based on a template with a homologous sequence and is used to precisely correct mutations or insert genes in a sophisticated repair mechanism (Chapman et al., 2012; Hsu et al., 2014; Kim and Kim, 2014; Liang et al., 1998).

In fact, genome engineering tools are used in a variety of fields such as genetic variation at the cellular level, making animal models, biofuel using algae, crop improvements as well as in the field of drug development and gene therapy (Hsu et al., 2014; Shao et al., 2016; Zhang et al., 2018). Recently, clinical trials for cancer T cell therapy using CRISPR have been under way (Rodriguez-Rodriguez et al., 2019).

Cas9 cleaves DNA double strands with two nuclease domains, two of which are structurally RuvC and HNH domains (Nishimasu et al., 2014). The 10th aspartic acid (D10), which is an active site in the RuvC domain, cuts non-target

strand that are not complementary to sgRNA, and 840 histidine (H840) is active site of HNH domain playing role of cutting target strand binding sgRNA. Alanine substitution for each of these active residues (D10A or H840A) can no longer cut double-stranded DNA but works as a nickase that cuts only one strand (Jinek et al., 2012; Nishimasu et al., 2014). These nickases can be used like nuclease by acting together as a paired nickase (Cho et al., 2014; Gopalappa et al., 2018; Ran et al., 2013). When used in combination with a repair template, it is also used to improve HR efficiency without NHEJ-caused indels compared to wild type Cas9 nuclease (Cullot et al., 2019; Satomura et al., 2017; Shao et al., 2018). Recently, it acts with the cytosine deaminase domain or the adenine deaminase domain, as a base editor. Dead Cas9, a catalytically impaired form (D10A H840A) that breaks down two active residues, can no longer cut DNA but still has the ability to bind to DNA, so it can be applied to a variety of applications. Transcription can be regulated by blocking transcription by binding to the region of the transcription start site (Vigouroux et al., 2018), or by acting fused with the transcription repressor domain (Kruppel associated box (KRAB)) (Yeo et al., 2018), activator (VP64, p65) (Chavez et al., 2015; Gilbert et al., 2014)). In addition, it can be fused with DNA demethylase (TET) (Xu et al., 2016), histone acetyl transferase (P300) (Hilton et al., 2015), and deacetylase (HDAC) (Kwon et al., 2017) to enable epigenetic control, or with fluorescent proteins for labeling chromosome (Ma et al., 2016a) as well as with nuclear compartment-specific proteins for 3D genome re-organization (Wang et al., 2018a). Recently, it has been applied to researches such as biosensors (Koo et al., 2018; Lee et al., 2018a).

One of the biggest limitations in the study using Cas9 is the low efficiency of HR. Since more than 50% of the mutations that can cause human genetic disease are point mutations (Rees and Liu, 2018), it is not enough to elaborate mutation to make a disease model or even to perform gene correction using Cas9 nucleases because HR efficiency is very low. The system that has been developed to supplement this is the base editor system which has been recently studied and which has received the most attention. In 2016, various cytosine base editors (CBEs) were reported and applied to several organisms; cytidine deaminases fused to Cas9 nickase or dead Cas9 (Gehrke et al., 2018; Hess et al., 2016; Jiang et al., 2018; Kim et al., 2017b; Komor et al., 2016; Li et al., 2018b; Ma et al., 2016b; Nishida et al., 2016; Wang et al., 2018b). The cytidine deaminases used were human-derived activation-induced cytidine deaminase (hAID), sea lamprey-derived PmCDA1, rat APOBEC1, and human-derived APOBEC3, which act on single stranded DNA (ssDNA) as well as single stranded RNA (ssRNA). These deaminases could target the C of non-target strand in R-loop formation. CBEs deaminate C, convert it to U, recognize it as T through base excision repair of the cell, and consequently convert C to T.

While CBEs almost directly fused existing cytidine deaminases with Cas9, adenosine deaminases in adenine base editors (ABEs) system had to be changed through directed evolution (Gaudelli et al., 2017). Many types of cytidine deaminases act on ssDNA, but adenosine deaminases are not yet known to function effectively on ssDNA. To this end, TadA, the tRNA adenosine deaminases of *Escherichia coli*, was engineered to act on ssDNA. When TadA was fused to Cas9

dead form and it was functioned on ssDNA in *E. coli*, selection was made by screening system to preserve spectinomycin resistant colonies. Because TadA acts as dimer in general (Losey et al., 2006), TadA wild type, TadA mutant containing 14 mutations produced through 7 rounds of directed evolution and Cas9 nickases fusion form was called ABE7.10. ABE7.10 was found to work well at various target sites without specific motifs at the human cell lines as well as in *E. coli*. The Cas9 nickases of ABEs nicks the target strand of the R-loop and the TadA heterodimer deaminates A of the non-target strand to inosine (I). During the cell repair, inosine is recognized as G by polymerase and thus A to G conversion occurs. ABEs have also been reported to be applicable not only to the cell level but also to various animals, plants and so on (Kang et al., 2018; Li et al., 2018a; Liu et al., 2018a; Liu et al., 2018b; Ryu et al., 2018).

An important issue when new CRISPR-based systems are developed is off-target effects. In order to apply to crop or livestock improvement or to see the therapeutic effect, it is necessary to identify and improve each Cas9 and sgRNA-based off-targets. Many methods have been devised to identify the genome-wide off-targets of Cas9 nuclease: an integrase-deficient lentiviral vector (IDLV) capture (Wang et al., 2015), high-throughput, genome-wide translocation sequencing (HTGTS) (Frock et al., 2015), direct in situ breaks labeling, enrichment on streptavidin, sequencing (BLESS) (Crosetto et al., 2013), breaks labeling in situ and sequencing (BLISS) (Yan et al., 2017), genome-wide, unbiased identification of DSBs evaluated by sequencing (GUIDE-seq) (Tsai et al., 2015), in vitro reporting of cleavage effects by sequencing (CIRCLE-seq) (Tsai et al., 2017),

selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-seq) (Cameron et al., 2017), and digested genome sequencing (Digenome-seq) (Kim et al., 2015). It is necessary to find a new off-target based on the fact that the base editors use a nickase rather than a nuclease and the efficiency varies depending on the position in the window of C or A. In the case of cytosine base editors, Digenome-seq was modified (Kim et al., 2017a). In vitro, genomic DNA was treated with CBE protein and one strand was cut and the other strand was cut with Uracil-Specific Excision Reagent (USER) enzyme (a mixture of *E. coli* Uracil DNA glycosylase (UDG) and DNA glycosylase-lyase Endonuclease VIII (Endo VIII)), which recognizes U, the deamination product of C, and aligned by whole genome sequencing potential off-target sites.

Here, I sought to profile ABE off-target activity in the human genome (Kim et al., 2019). First of all, I increased substitution efficiency of ABEs by optimizing codon usage and increasing the number of nuclear localization signals. Then, I present a modified version of Digenome-seq to assess genome-wide target specificity of ABEs. To produce DSBs at sites containing inosines, the products of adenine deamination, human genomic DNA is treated with an ABE7.10 protein-guide RNA complex and either Endonuclease V (Endo V) or a combination of human Alkyladenine DNA Glycosylase (hAAG) and Endo VIII *in vitro*. Digenome-seq detects ABE off-target sites with a substitution frequency of 0.1% or more. As expected, ABE7.10, and BE3, a cytosine base editor, and unmodified Cas9 often recognize different off-target sites, highlighting the need for independent assessments of their genome-wide specificities (Kim et al., 2017a).

Using targeted sequencing, we also show that use of preassembled ABE ribonucleoproteins, modified guide RNAs (Cho et al., 2014; Fu et al., 2014; Kim et al., 2015; Kim et al., 2016; Tsai et al., 2015), and Sniper-Cas9 (Lee et al., 2018b) reduces ABE off-target activity and increases specificities in human cells.

Ⅱ. Materials and Methods

1. Cell culture and transfection

HEK293T cells (ATCC CRL-11268) were cultured in DMEM medium supplemented with 10% FBS and 1% penicillin/streptomycin (Welgene) and verified using STR profiling.

HEK293T cells (0.8 x 10^5) were seeded into 48-well plates 24 hours before transfection and transfected at 70–80% confluency with plasmids encoding ABE (750 ng of plasmid DNA), BE3 (750 ng), or Cas9 (750 ng) together with a plasmid encoding sgRNA (250 ng) using Lipofectamine 2000 (Invitrogen; 1.5 μL for each well).

For ABE RNP-mediated genome editing, HEK293T cells (1 x 10^5) were electroporated with ABE7.10 protein (15 μg) and in vitro transcribed sgRNA (7.5 μg) or plasmids encoding ABE (0.6 μg) and sgRNA (0.2 μg) via a 10 μl Neon Transfection System with program number 15.

Cells were lysed with lysis buffer (50 mM Tris-HCl, 1 mM EDTA, 0.005% SDS, pH 8.0) and DNA was amplified and analyzed after 72 hours post-transfection.

2. Expression and purification of ABE7.10 protein

A plasmid encoding the human codon-optimized ABE7.10 with an N-terminal His purification tag was transformed into BL21 Star (DE3)-competent *E.*

coli cells. Following transformation, a fresh single colony was grown overnight in Luria-Bertani (LB) broth containing 50 μg/mL kanamycin with shaking at 37℃. 8 mL pre-cultures diluted at 1:50 were inoculated into individual flasks containing 400 mL of LB broth supplemented with 50 μg/mL kanamycin; these cultures were incubated with shaking at 37℃ until OD_{600}=0.65~0.70; thereafter, the cultures were kept on ice. Protein expression was induced with 1 mM isopropyl-b-D-1-thiogalactopyranoside (IPTG; GoldBio), after which cultures were incubated for 16 hours with shaking at 18℃.

The subsequent protein purification procedures were carried out at 4℃. Cells were harvested by centrifugation at 6,000 xg for 10 minutes, and re-suspended in lysis buffer (50 mM sodium phosphate (Sigma-Aldrich), 500 mM NaCl (Sigma-Aldrich), 10 mM imidazole (Sigma-Aldrich), 1% Triton x-100 (Sigma-Aldrich), 20% glycerol, 1 mM 1,4-dithiothreitol (DTT; GoldBio), 1 mM phenylmethylsulfonyl fluoride (PMSF; Sigma-Aldrich), concentration of 1 mg/mL lysozyme (Sigma-Aldrich), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). The suspensions were then alternatively incubated in liquid nitrogen and a water bath (37℃), repeated 3 times. Cells were lysed by sonication for 9 minutes (5 seconds (on), 10 seconds (off)), and the lysates were cleared by centrifugation at 15,000 xg for 20 minutes. The supernatant was incubated with nickel agarose beads (Ni-NTA, QIAGEN) for 60 minutes with shaking at 4℃. The lysate-resin mixture was loaded into a polypropylene column, washed 2 times with 3 column volumes of wash buffer (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 35 mM imidazole (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM

$ZnCl_2$ (Sigma-Aldrich), pH 8.0), and bound protein was eluted with elution buffer (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 250 mM imidazole (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). The eluted protein fraction was next loaded into a polypropylene column that contained heparin agarose beads (Heparin Sepharose 6 Fast Flow, GE Healthcare) and washed 2 times with 3 column volumes of wash buffer (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). Bound protein was eluted with elution buffer (50 mM sodium phosphate (Sigma-Aldrich), 750 mM NaCl (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). The resulting protein fractions ware concentrated with a centrifuge column (Amicon Ultra-4 Contrifugal Filter Devices, Millipore) at 6,000 xg.

3. *In vitro* transcription of sgRNAs

For a 100 μL in vitro transcription reaction, sgRNA templates containing T7 promoter sequence were incubated with each 4 mM NTP (rATP, rGTP, rCTP, rUTP), additional 14 mM $MgCl_2$ and 100 units of RNase inhibitor (New England BioLabs), 500 units of T7 RNA polymerase (New England BioLabs) in T7 RNA polymerase buffer for 3 hours at 37℃. Subsequently, the reaction mixture was treated with DNase I (New England BioLabs) for 30 minutes at 37℃ and the in vitro-transcribed gRNAs were purified using an QIAGEN miRNeasy kit according to the manufacturer's protocol.

4. ABE7.10-mediated in vitro digestion of a PCR amplicon

A PCR amplicon containing the *HEK2* site was incubated with the ABE7.10 protein and in vitro transcribed *HEK2*-targeting sgRNA in a reaction volume of 100 μL for 1 hour at 37 °C. The deaminased product was then purified using a PCR purification kit (MGmed). 2 μg of deaminased product was incubated with Endonuclease V (40 units) (New England BioLabs) or human Alkyladenine DNA Glycosylase (10 units) (New England BioLabs) with Endonuclease VIII (20 units) (New England BioLabs) in a reaction volume of 200 μL for 30 minutes at 37 °C, after which the DNA was purified again using a PCR purification kit (MGmed) and subjected to analysis with a Bioanalyzer (Agilent) to determine the results of the in vitro digestion.

5. ABE7.10-mediated in vitro digestion of genomic DNA.

Genomic DNA was isolated from HEK293T cells using a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. 8 μg of genomic DNA was incubated with the purified ABE7.10 protein (300 nM) and an sgRNA (900 nM) in a reaction volume of 400 μL in reaction buffer (50 mM Tris-HCl (pH 8.0), 25 mM KCl, 2.5 mM $MgSO_4$, 0.1 mM EDTA, 10% glycerol, 2 mM DTT, 10 μM $ZnCl_2$) at 37 °C for 8 hours. The DNA was then purified with a DNeasy Blood & Tissue Kit (Qiagen) after removal of sgRNA using RNase A (50 μg/mL). 3 μg of purified DNA was incubated with Endonuclease V (40 units) or human Alkyladenine DNA Glycosylase (10 units) with Endonuclease VIII (20 units) in a reaction volume of 200 μL at 37 °C for 2 hours, then was purified again with a

DNeasy Blood & Tissue Kit (Qiagen). Purified DNA was incubated with KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) and analyzed by real-time quantitative PCR (qPCR) to check for ABE7.10- and Endonuclease V- or ABE7.10-, human Alkyladenine DNA Glycosylase-, and Endonuclease VIII-mediated DNA DSBs. The fraction of intact genomic DNA was measured using the comparative C_T method (Schmittgen and Livak, 2008).

6. Whole genome sequencing and Digenome sequencing

1 μg of in vitro digested DNA was fragmented to the 400- to 500-bp range using the Covaris system (Thermo Fisher Scientific) and incubated with End Repair Mix (Illumina) to generate blunt ends. Fragmented DNA was ligated with adapters to produce libraries and subjected to WGS using a HiSeq X Ten Sequencer (Illumina) at Macrogen. WGS was performed at a sequencing depth of 30–40×. Isaac aligner was used to map the genome sequence. DNA cleavage sites were identified using Digenome 1.0 and Digenome 2.0 programs. The source code of the version of Digenome 2.0 used in this manuscript is available at https://github.com/chizksh/digenome-toolkit2.

7. Targeted deep sequencing

On- and potential off-target sites were amplified using KAPA HiFi HotStart PCR polymerase(#KK2502). Amplicons were again amplified using TruSeq HT Dual index-containing primers to generate deep sequencing libraries. The libraries were sequenced using Illumina MiniSeq with paired-end sequencing

systems. Base editing frequencies indicate the frequencies of modified target sites with at least one edit within the editing window (position 4-7). The source code of the computer program to calculate substitution and indel frequencies are available at https://github.com/ibs-cge/maund.

8. Statistical analysis

All statistical results are means ± s.e.m. from three independent biological replicates. The significance of the differences between the two other groups was calculated using the two-tailed student's t-test method. One asterisk (*) indicates $P < 0.05$ in the figures and figure legends. Data were analyzed with Graph Pad PRISM 7 and Microsoft Excel.

Ⅲ. Results

1. Base editing efficiency according to the codon usage and the number of nuclear localization signals

ABEs consist of mutant Cas9 nickases and mutant *E. coli* TadA modified to deaminate DNA through directed evolution (Gaudelli et al., 2017). Therefore, various variants exist depending on the evolution stage. For example, there are ABE6.3, ABE7.8, ABE7.9, ABE7.10. First, I confirmed the substitution efficiency of these variants in two cell lines, HeLa and HEK293T. I cloned each TadA variant into Cas9 expressing plasmid vector, which we already have in our group, and compared it with the plasmids used in the reported paper (Gaudelli et al., 2017) and deposited on Addgene. Previous reports have shown that Cas9 expressing plasmid of our group has higher expression levels and induces higher mutation frequency due to codon usage than Cas9 plasmids in other groups (Kim et al., 2017c). In fact, we found that the efficiency of the variant construct based on Cas9 of our group was higher in *HEK2*, *HEK3* and *RNF2* target sites generally (Figure 1). In particular, the effect of the CMV promoter expressing the base editor was relatively weak in the HeLa cell line, so the effect was more dramatic (Mao et al., 2008). This result is in agreement with the previous paper that the expression level depending on the codon of the DNA encoding the base editors can affect the mutation frequency (Koblan et al., 2018; Zafra et al., 2018).

Among the variants of ABEs, the most widely used variant is ABE7.10,

which is the most efficient variant. There has been a report on ABEmax, a construct that has newly synthesized DNA encoding ABE7.10 and used a new codon and switched nuclear localization signals (NLSs) from one SV40 NLS to two Bipartite NLS (BpNLS) (Koblan et al., 2018). ABE (Broad) deposited in the Addgene, ABE (SNU), ABEmax (Broad), and ABEmax (SNU) constructs, which replaces one SV40NLS of ABE (SNU) with two BpNLS, were compared (Figure 2a). Comparing the efficiency at 23 endogenous target sites, the base editing frequency was increased when the type and number of NLS were changed in both HEK293T cell line and HeLa cell line (Figure 2b-c, Table 1-2). The efficiency of ABEmax (SNU) was the highest in the HEK293T cell line (Figure 2b) and the ABEmax (Broad) efficiency was the highest in the HeLa cell line. Based on these results, it was confirmed that the codon usage and the number and type of NLS have a significant effect on the efficiency.

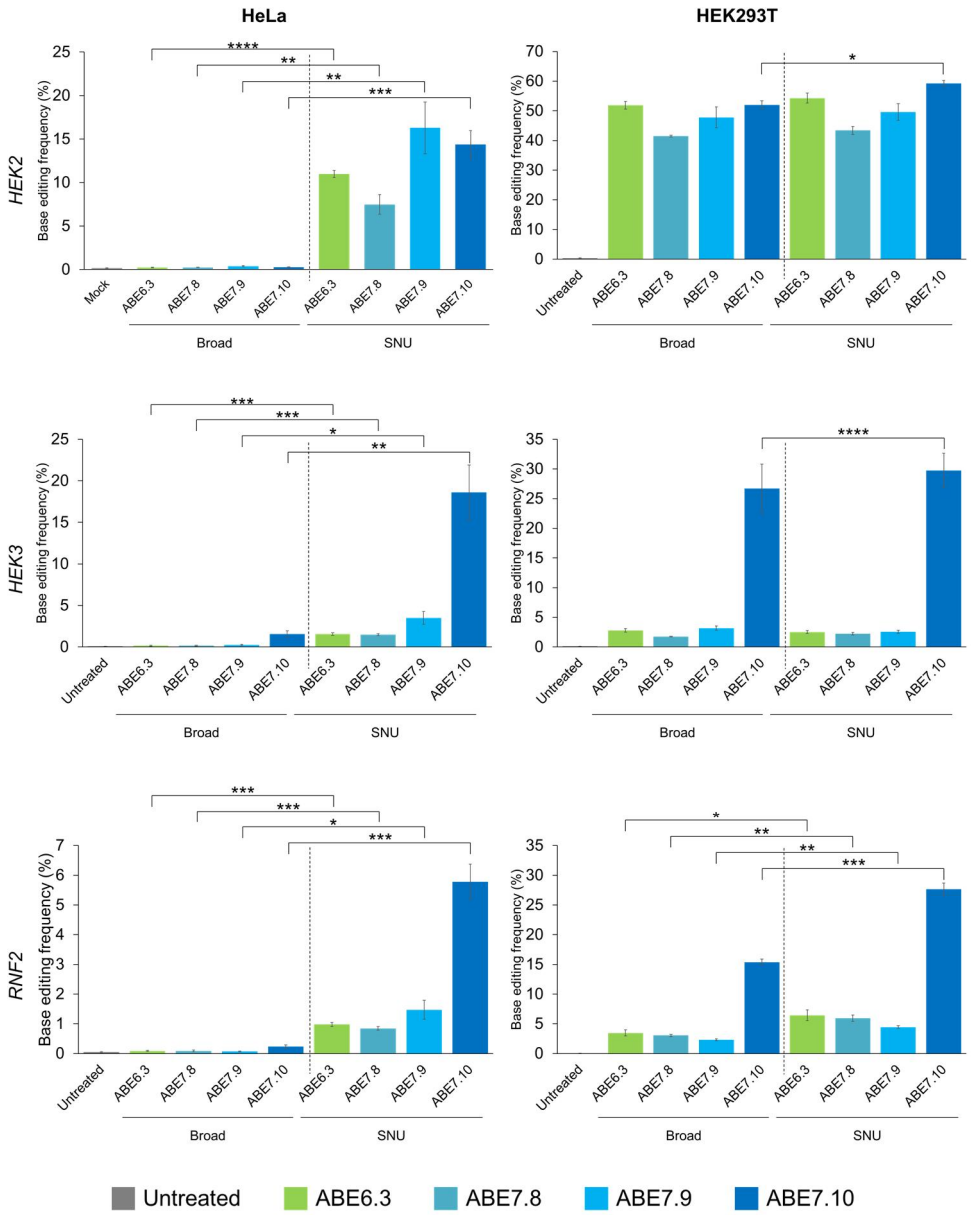


Figure 1. Comparison of substitution frequency using several versions of adenine base editors with different codon usages in HeLa and HEK293T cell line. Base editing efficiencies were measured by targeted deep sequencing. Means

± s.e.m. were from three independent experiments. Student's t-test (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****: $p<0.0001$)

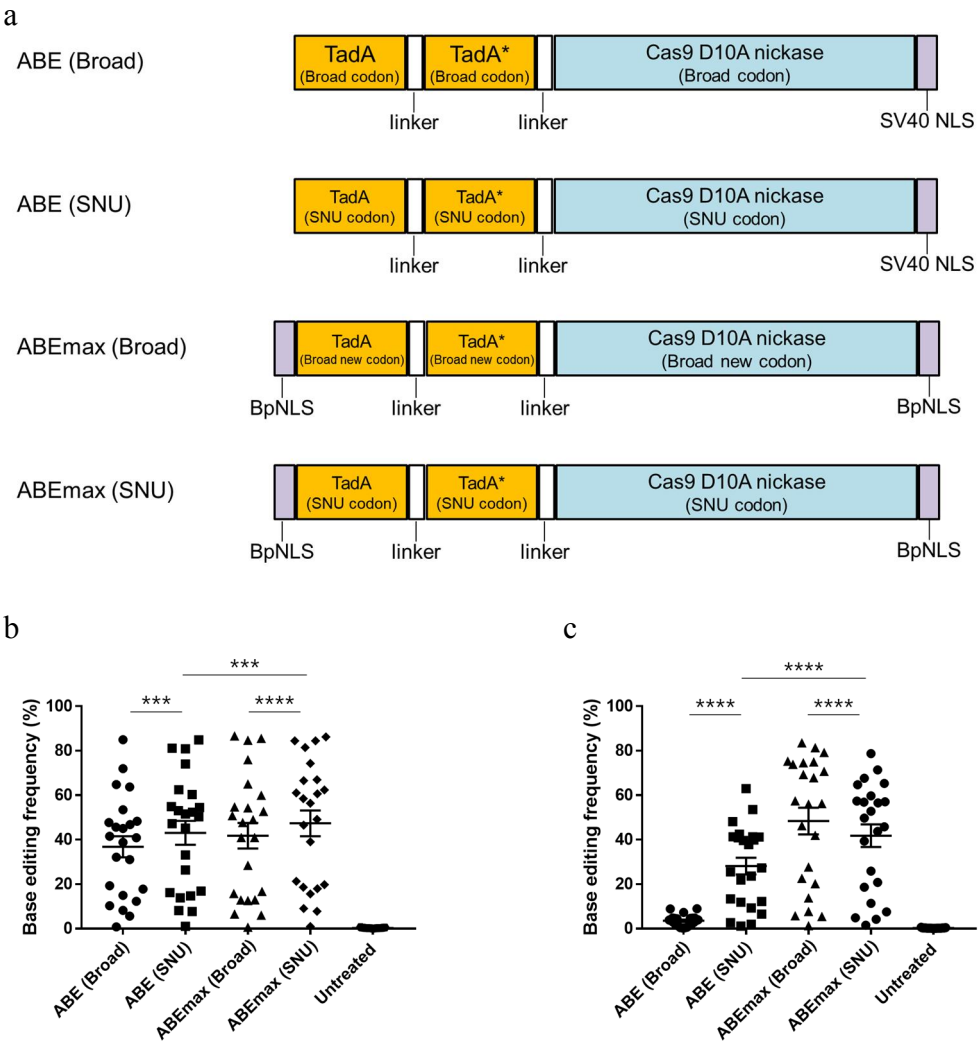


Figure 2. Comparison of base editing frequency according to codon usages and the number of NLSs in HEK293T and HeLa cell line. (**a**) Plasmid constructs for testing the efficiency dependent on the codon usage and the number of NLS. ABE and ABEmax are distinguished by the type and number of NLS. ABE contains only one SV40 NLS at C-terminal, and ABEmax contains two bpNLS at both the N and C termini. **b**, **c**, Base editing frequencies using 4 different ABE

constructs in HEK293T cell lines (**b**) and HeLa cell line (**c**). Total 23 target sites. Each dot represents the mean value of the results of the three independent experiments. Base editing efficiencies were measured by targeted deep sequencing. Error bars indicate s.e.m. Paired student's t-test (***: $p<0.001$, ****: $p<0.0001$)

Table 1. Base editing frequencies of ABE7.10 depending on codon usage of vector constructs in HEK293T cell lines

			Base editing frequency (%)				
	Gene	Target sequence	ABE (Broad)	ABE (SNU)	ABEmax (Broad)	ABEmax (SNU)	Untreated
1	FANCF	GGAATCCCTTCTGCAGCACCTGG	10.3	14.7	15.8	19.8	0.2
2	RNF2	GTCATCTTAGTCATTACCTGAGG	32.1	45.2	40.9	46.5	0.2
3	HEK1	GGGAAAGACCCAGCATCCGTGGG	17.8	16.9	12.9	17.9	0.3
4	HEK2	GAACACAAAGCATAGACTGCGGG	84.9	84.9	86.6	84.4	0.7
5	HEK3	GGCCCAGACTGAGCACGTGATGG	41.6	33.1	28.5	39.0	0.2
6	HEK4	GGCACTGCGGCTGGAGGTGGGGG	8.2	7.7	6.5	7.8	0.0
7	HPRT_E2_1	GAAAGGGTGTTTATTCCTCATGG	5.6	8.1	6.0	9.0	0.2
8	HPRT_E3_1	GATGTGATGAAGGAGATGGGAGG	12.3	13.9	12.6	15.6	0.1
9	HPRT_E4_1	GGGGACATAAAAGTAATTGGTGG	45.0	51.6	54.3	61.0	0.1
10	HPRT_E4_2	GATGATCTCTCAACTTTAACTGG	19.3	26.4	16.7	21.3	0.2
11	HPRT_E6_1	GTATAATCCAAAGATGGTCAAGG	46.8	62.5	60.0	66.9	0.3
12	HPRT_E8_1	GAAGTATTCATTATAGTCAAGGG	72.0	80.8	84.7	84.5	0.3
13	TK_CTK	GCGGAAACACGGGACCAAGTCGG	15.0	16.2	12.9	18.6	0.5
14	TK_EphA1	GCTCCAATTGGATCTACCGCGGG	48.3	60.3	65.1	74.3	0.3
15	TK_EphB4	GCAGAATATTCGGACAAACACGG	40.9	53.1	47.7	56.4	0.3
16	TK_INSR	GAGAATTGCTCTGTCATCGAAGG	31.1	54.4	54.7	66.5	0.3
17	TK_ROR1	GCCATAGATGGTGGACCGAAAGG	53.4	52.4	52.7	58.5	0.4
18	TK_TYRO3	GGCCACACTAGCGTTGCTGCTGG	64.8	81.1	85.6	86.2	0.5
19	TK_WEE1	GTCGAGATGTTCTATTACTCTGG	45.6	54.8	49.0	60.7	0.2
20	CCR5-3	GGCAGCATAGTGAGCCCAGAAGG	63.7	74.0	76.0	81.5	0.3
21	CCR5-10	GGTGACAAGTGTGATCACTTGGG	47.7	47.2	41.0	49.1	0.4
22	CCR5-15	GACAAGTGTGATCACTTGGGTGG	38.8	50.3	50.8	62.4	0.4
23	CCR5-17	GACACCGAAGCAGAGTTTTTAGG	0.8	1.1	0.6	1.0	0.4

Table 2. Base editing frequencies of ABE7.10 depending on codon usage of vector constructs in HeLa cell lines

			Base editing frequency (%)				
	Gene	Target sequence	ABE (Broad)	ABE (SNU)	ABEmax (Broad)	ABEmax (SNU)	Untreated
1	FANCF	GGAATCCCTTCTGCAGCACCTGG	2.0	12.2	22.6	18.6	0.2
2	RNF2	GTCATCTTAGTCATTACCTGAGG	0.7	13.4	27.7	25.9	0.2
3	HEK1	GGGAAAGACCCAGCATCCGTGGG	1.2	9.3	13.8	11.4	0.3
4	HEK2	GAACACAAAGCATAGACTGCGGG	4.0	39.7	70.7	56.6	0.6
5	HEK3	GGCCCAGACTGAGCACGTGATGG	9.0	41.2	56.0	52.8	0.2
6	HEK4	GGCACTGCGGCTGGAGGTGGGGG	1.3	6.5	7.7	7.5	0.1
7	HPRT_E2_1	GAAAGGGTGTTTATTCCTCATGG	0.4	1.9	5.5	4.2	0.2
8	HPRT_E3_1	GATGTGATGAAGGAGATGGGAGG	0.3	2.7	5.8	4.9	0.1
9	HPRT_E4_1	GGGGACATAAAAGTAATTGGTGG	4.8	37.8	69.3	57.0	0.1
10	HPRT_E4_2	GATGATCTCTCAACTTTAACTGG	2.7	21.9	41.9	39.4	0.2
11	HPRT_E6_1	GTATAATCCAAAGATGGTCAAGG	3.9	41.3	73.9	56.8	0.2
12	HPRT_E8_1	GAAGTATTCATTATAGTCAAGGG	7.3	63.0	79.2	78.7	0.2
13	TK_CTK	GCGGAAACACGGGACCAAGTCGG	1.6	11.9	20.2	20.8	0.5
14	TK_EphA1	GCTCCAATTGGATCTACCGCGGG	4.7	41.3	74.5	65.3	0.2
15	TK_EphB4	GCAGAATATTCGGACAAACACGG	4.5	42.5	74.9	64.7	0.2
16	TK_INSR	GAGAATTGCTCTGTCATCGAAGG	2.7	27.3	57.5	43.7	0.2
17	TK_ROR1	GCCATAGATGGTGGACCGAAAGG	4.6	39.9	75.3	57.4	0.3
18	TK_TYRO3	GGCCACACTAGCGTTGCTGCTGG	4.6	48.1	83.6	67.6	0.4
19	TK_WEE1	GTCGAGATGTTCTATTACTCTGG	3.5	40.9	67.8	59.7	0.1
20	CCR5-3	GGCAGCATAGTGAGCCCAGAAGG	8.9	53.5	81.4	71.4	0.3
21	CCR5-10	GGTGACAAGTGTGATCACTTGGG	4.8	25.6	46.3	45.8	0.3
22	CCR5-15	GACAAGTGTGATCACTTGGGTGG	3.5	23.6	56.2	49.7	0.3
23	CCR5-17	GACACCGAAGCAGAGTTTTTAGG	0.4	1.1	1.3	1.5	0.4

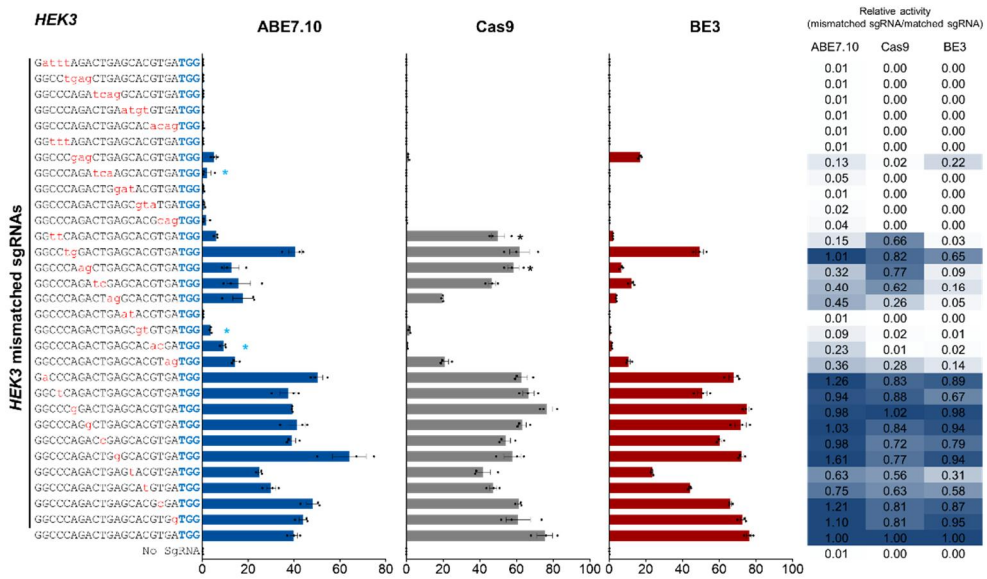
2. Tolerance of ABE, BE3 and Cas9 for mismatched sgRNAs

a. Mismatch tolerance of ABE7.10 is independent of Cas9 nucleases and BE3 cytosine base editors

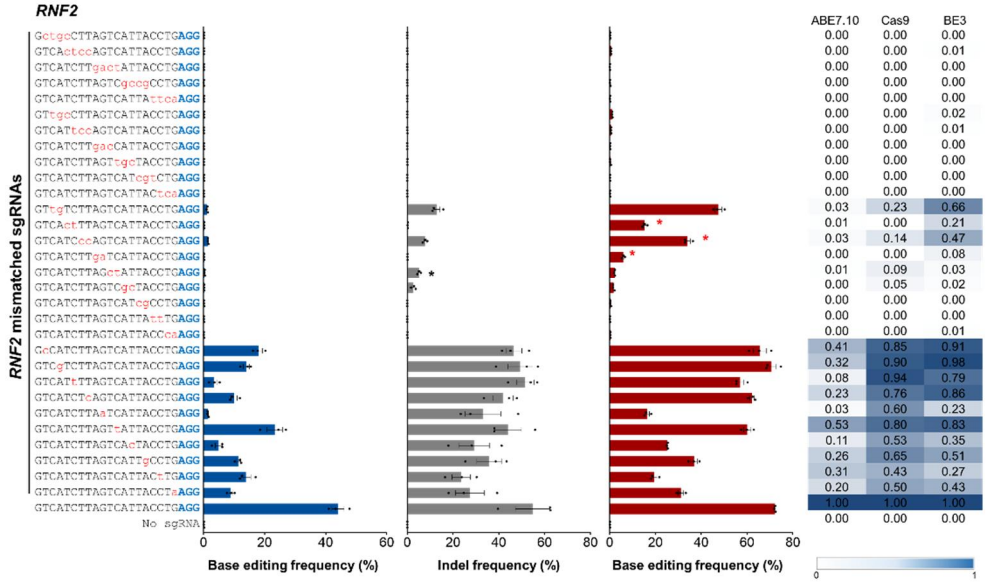
I first investigated whether off-target activities of ABE7.10, BE3 (a cytosine base editor), and Cas9 are different from each other by testing a series of mismatched sgRNAs targeted to endogenous genomic loci (Figure 3a-c and Table 3). To this end, I transfected HEK293T cells with plasmids encoding sgRNAs having one- to four-nucleotide mismatches and plasmids encoding human codon-optimized ABE7.10, BE3, or SpCas9 and measured frequencies of single-nucleotide substitutions or small insertions and deletions (indels) using high-throughput DNA sequencing. As expected, base editors and Cas9 nucleases tolerated most of the single or double mismatches in the sgRNAs but were poorly active when combined with most of the sgRNAs having triple or quadruple mismatches. Notably, the tolerance of Cas9, ABE7.10, and BE3 for mismatched sgRNAs was often different from each other. For example, at the *HEK3* site, Cas9 complexed with an sgRNA having tandem mismatches at positions 3-4 (numbered 1 to 23 in the 5' to 3' direction) was highly active, showing a relative activity of 66% compared to the fully-matched sgRNA, whereas ABE7.10 complexed with the same sgRNA was poorly active, showing a relative activity of 15% (Figure 3a). At the *HEK2* site, several sgRNAs with 2~4 mismatches (indicated with asterisks in Figure 3c) were well-tolerated by ABE7.10 but were poorly active or inactive with Cas9 or BE3. Also, the overall correlation between ABE-mediated substitution

frequency and Cas9-mediated indel frequency or BE3-mediated substitution frequency was not always high, depending on the target site (Figure 4). These results suggest that ABE7.10, BE3, and Cas9 could recognize separate sets of off-target sites in the human genome, calling for a method to determine genome-wide ABE specificity in an unbiased manner.

a



b



(*Continued*)

c

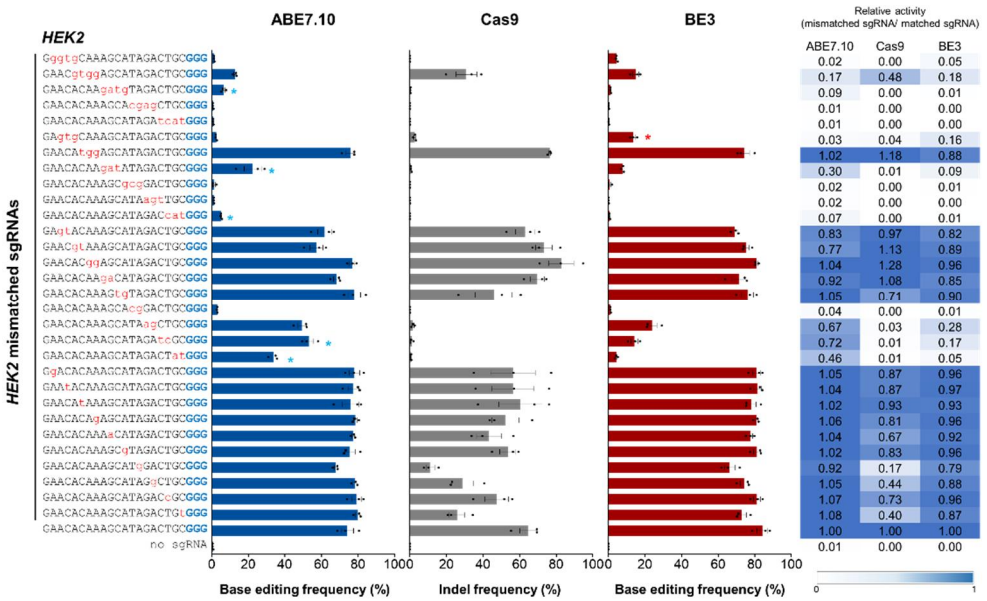


Figure 3. Tolerance of ABE7.10, BE3, and Cas9 for mismatched sgRNAs. Mismatched sgRNAs that differed from the *HEK3* site (a), *RNF2* (b), and *HEK2* site (c) by one to four nucleotides were tested in HEK293T cells. Base editing or indel frequencies were measured using targeted deep sequencing. Mismatched nucleotides and the PAM sequence are shown in red and blue, respectively. Asterisks indicate mismatched sgRNAs whose relative activity (mismatched sgRNA/matched sgRNA) with one enzyme is more than three times higher than that with the other two enzymes. Means ± s.e.m. were from three independent experiments.

Table 3. Mutation frequencies of ABE7.10, Cas9, and BE3 at HEK3, RNF2, and HEK2 sites using mismatched sgRNAs

	HEK3			
		Base editing frequency (%)	Indel frequency (%)	Base editing frequency (%)
		(+) ABE7.10	(+) Cas9	(+) BE3
HEK3 mismatched sgRNAs	GatttAGACTGAGCACGTGATGG	0.32	0.00	0.17
	GGCCtgagCTGAGCACGTGATGG	0.41	0.03	0.13
	GGCCCAGAtcagGCACGTGATGG	0.48	0.05	0.20
	GGCCCAGACTGAatgtGTGATGG	0.31	0.09	0.12
	GGCCCAGACTGAGCACacagTGG	0.39	0.02	0.25
	GGtttAGACTGAGCACGTGATGG	0.33	0.02	0.15
	GGCCCgagCTGAGCACGTGATGG	5.23	1.15	16.86
	GGCCCAGAtcaAGCACGTGATGG	2.18	0.00	0.21
	GGCCCAGACTGgatACGTGATGG	0.56	0.01	0.27
	GGCCCAGACTGAGCgtaTGATGG	0.79	0.01	0.18
	GGCCCAGACTGAGCACGcagTGG	1.54	0.18	0.22
	GGttCAGACTGAGCACGTGATGG	6.02	49.87	2.11
	GGCCtgGACTGAGCACGTGATGG	40.39	61.63	49.37
	GGCCCAagCTGAGCACGTGATGG	12.81	58.23	6.68
	GGCCCAGAtcGAGCACGTGATGG	15.88	46.57	12.28
	GGCCCAGACTagGCACGTGATGG	17.80	19.65	3.98
	GGCCCAGACTGAatACGTGATGG	0.46	0.04	0.19
	GGCCCAGACTGAGCgtGTGATGG	3.63	1.66	0.64
	GGCCCAGACTGAGCACacGATGG	9.28	0.51	1.53
	GGCCCAGACTGAGCACGTagTGG	14.31	21.05	10.61
	GaCCCAGACTGAGCACGTGATGG	50.32	62.84	67.85
	GGCtCAGACTGAGCACGTGATGG	37.32	66.54	50.80
	GGCCCgGACTGAGCACGTGATGG	39.21	76.56	74.92
	GGCCCAGgCTGAGCACGTGATGG	41.09	63.21	71.77
	GGCCCAGACcGAGCACGTGATGG	38.99	53.99	60.41
	GGCCCAGACTGgGCACGTGATGG	64.05	57.72	72.00
	GGCCCAGACTGAGtACGTGATGG	25.02	41.86	23.39
	GGCCCAGACTGAGCAtGTGATGG	30.04	47.29	43.94
	GGCCCAGACTGAGCACGcGATGG	48.08	61.45	66.11
	GGCCCAGACTGAGCACGTGgTGG	43.88	60.87	72.75
	GGCCCAGACTGAGCACGTGATGG	39.87	75.41	76.33
	no sgRNA	0.21	0.00	0.16

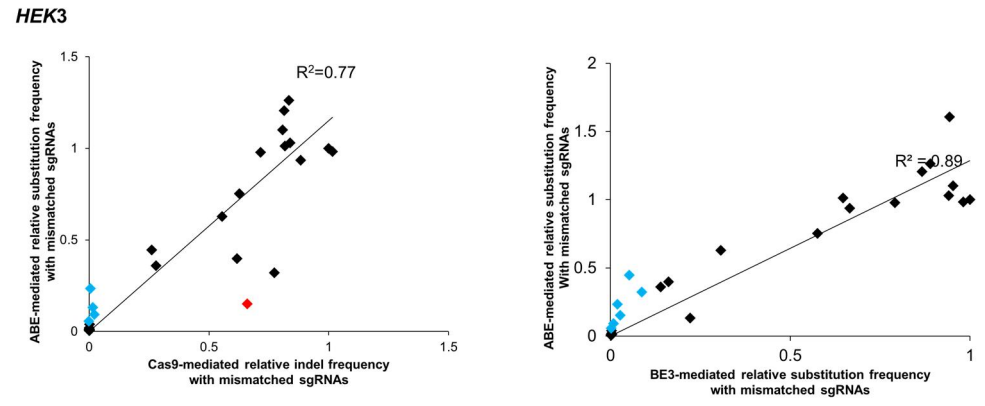
(Continued)

	RNF2			
		Base editing frequency (%)	Indel frequency (%)	Base editing frequency (%)
		(+) ABE7.10	(+) Cas9	(+) BE3
RNF2 mismatched sgRNAs	GctgcCTTAGTCATTACCTGAGG	0.13	0.01	0.27
	GTCActccAGTCATTACCTGAGG	0.12	0.04	0.60
	GTCATCTTgactATTACCTGAGG	0.12	0.01	0.11
	GTCATCTTAGTCgccgCCTGAGG	0.10	0.02	0.20
	GTCATCTTAGTCATTAttcaAGG	0.14	0.01	0.16
	GTtgcCTTAGTCATTACCTGAGG	0.15	0.01	1.13
	GTCATtccAGTCATTACCTGAGG	0.10	0.03	0.65
	GTCATCTTgacCATTACCTGAGG	0.12	0.00	0.23
	GTCATCTTAGTtgcTACCTGAGG	0.12	0.06	0.27
	GTCATCTTAGTCATcgtCTGAGG	0.14	0.01	0.16
	GTCATCTTAGTCATTACtcaAGG	0.10	0.01	0.20
	GTtgTCTTAGTCATTACCTGAGG	1.22	12.77	47.48
	GTCActTTAGTCATTACCTGAGG	0.22	0.23	15.33
	GTCATCccAGTCATTACCTGAGG	1.44	7.76	34.05
	GTCATCTTgaTCATTACCTGAGG	0.20	0.10	5.97
	GTCATCTTAGctATTACCTGAGG	0.38	5.16	2.19
	GTCATCTTAGTCgcTACCTGAGG	0.21	2.74	1.67
	GTCATCTTAGTCATcgCCTGAGG	0.17	0.02	0.33
	GTCATCTTAGTCATTAttTGAGG	0.14	0.02	0.13
	GTCATCTTAGTCATTACCcaAGG	0.12	0.08	0.37
	GcCATCTTAGTCATTACCTGAGG	18.06	46.51	65.55
	GTCgTCTTAGTCATTACCTGAGG	14.01	49.31	70.67
	GTCATtTTAGTCATTACCTGAGG	3.53	51.54	56.95
	GTCATCTcAGTCATTACCTGAGG	9.98	41.89	62.12
	GTCATCTTAaTCATTACCTGAGG	1.37	33.09	16.41
	GTCATCTTAGTtATTACCTGAGG	23.31	43.94	60.08
	GTCATCTTAGTCAcTACCTGAGG	4.99	29.11	25.20
	GTCATCTTAGTCATTgCCTGAGG	11.45	35.75	36.77
	GTCATCTTAGTCATTACtTGAGG	13.85	23.51	19.42
	GTCATCTTAGTCATTACCTaAGG	8.98	27.32	31.08
	GTCATCTTAGTCATTACCTGAGG	44.00	54.83	72.17
	no sgRNA	0.13	0.03	0.12

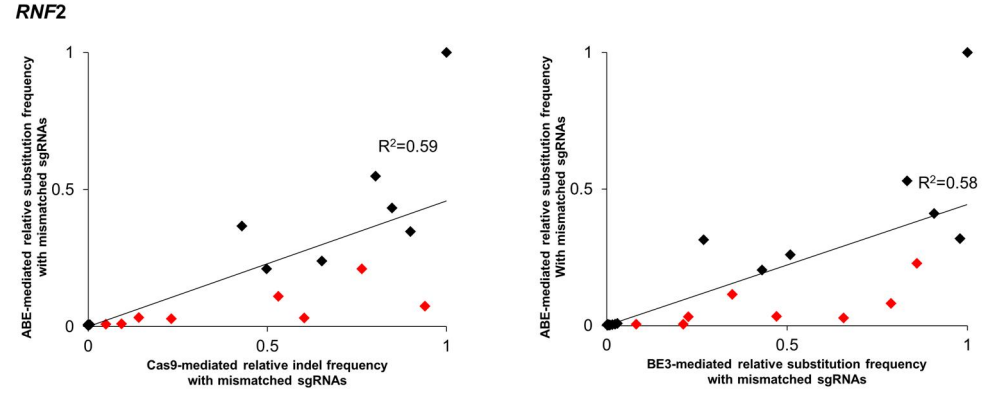
(Continued)

	HEK2			
		Base editing frequency (%)	Indel frequency (%)	Base editing frequency (%)
		(+) ABE7.10	(+) Cas9	(+) BE3
HEK2 mismatched sgRNAs	GggtgCAAAGCATAGACTGCGGG	1.33	0.03	4.53
	GAACgtggAGCATAGACTGCGGG	12.76	30.83	15.00
	GAACACAAgatgTAGACTGCGGG	6.48	0.30	1.18
	GAACACAAAGCAcgagCTGCGGG	0.70	0.00	0.18
	GAACACAAAGCATAGAtcatGGG	0.86	0.03	0.19
	GAgtgCAAAGCATAGACTGCGGG	2.40	2.60	13.68
	GAACAtggAGCATAGACTGCGGG	75.71	76.37	74.30
	GAACACAAgatATAGACTGCGGG	22.45	0.64	7.76
	GAACACAAAGCgcgGACTGCGGG	1.33	0.00	0.75
	GAACACAAAGCATAagtTGCGGG	1.16	0.05	0.19
	GAACACAAAGCATAGACcatGGG	5.12	0.02	0.52
	GAgtACAAAGCATAGACTGCGGG	61.71	63.07	69.18
	GAACgtAAAGCATAGACTGCGGG	57.23	73.27	75.35
	GAACACggAGCATAGACTGCGGG	76.76	82.69	80.95
	GAACACAAgaCATAGACTGCGGG	68.19	69.70	71.33
	GAACACAAAGtgTAGACTGCGGG	77.93	45.78	76.05
	GAACACAAAGCAcgGACTGCGGG	3.01	0.03	1.10
	GAACACAAAGCATAagCTGCGGG	49.41	1.68	23.86
	GAACACAAAGCATAGAtcGCGGG	53.12	0.96	14.18
	GAACACAAAGCATAGACTatGGG	33.85	0.54	4.27
	GgACACAAAGCATAGACTGCGGG	77.99	56.26	80.99
	GAAtACAAAGCATAGACTGCGGG	77.47	56.28	81.51
	GAACAtAAAGCATAGACTGCGGG	75.90	60.44	77.98
	GAACACAgAGCATAGACTGCGGG	78.61	52.17	80.94
	GAACACAAAaCATAGACTGCGGG	77.42	43.30	77.51
	GAACACAAAGCgTAGACTGCGGG	75.62	53.51	81.23
	GAACACAAAGCATgGACTGCGGG	67.88	11.14	66.25
	GAACACAAAGCATAGgCTGCGGG	78.22	28.65	74.33
	GAACACAAAGCATAGACcGCGGG	79.11	47.36	81.18
	GAACACAAAGCATAGACTGtGGG	79.86	25.81	73.01
	GAACACAAAGCATAGACTGCGGG	74.14	64.70	84.20
	no sgRNA	0.63	0.00	0.07

a


HEK3
R²=0.77
ABE-mediated relative substitution frequency with mismatched sgRNAs
Cas9-mediated relative indel frequency with mismatched sgRNAs
R² =0.89
ABE-mediated relative substitution frequency With mismatched sgRNAs
BE3-mediated relative substitution frequency with mismatched sgRNAs

b


RNF2
R²=0.59
ABE-mediated relative substitution frequency with mismatched sgRNAs
Cas9-mediated relative indel frequency with mismatched sgRNAs
R²=0.58
ABE-mediated relative substitution frequency With mismatched sgRNAs
BE3-mediated relative substitution frequency with mismatched sgRNAs

c

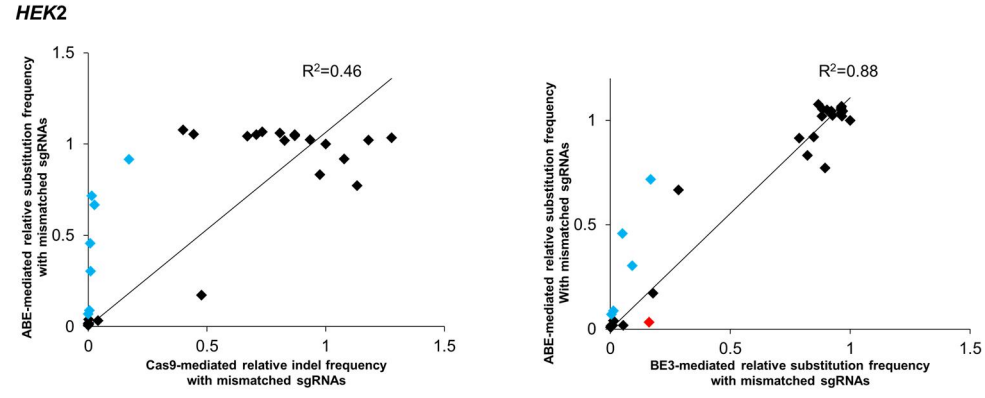
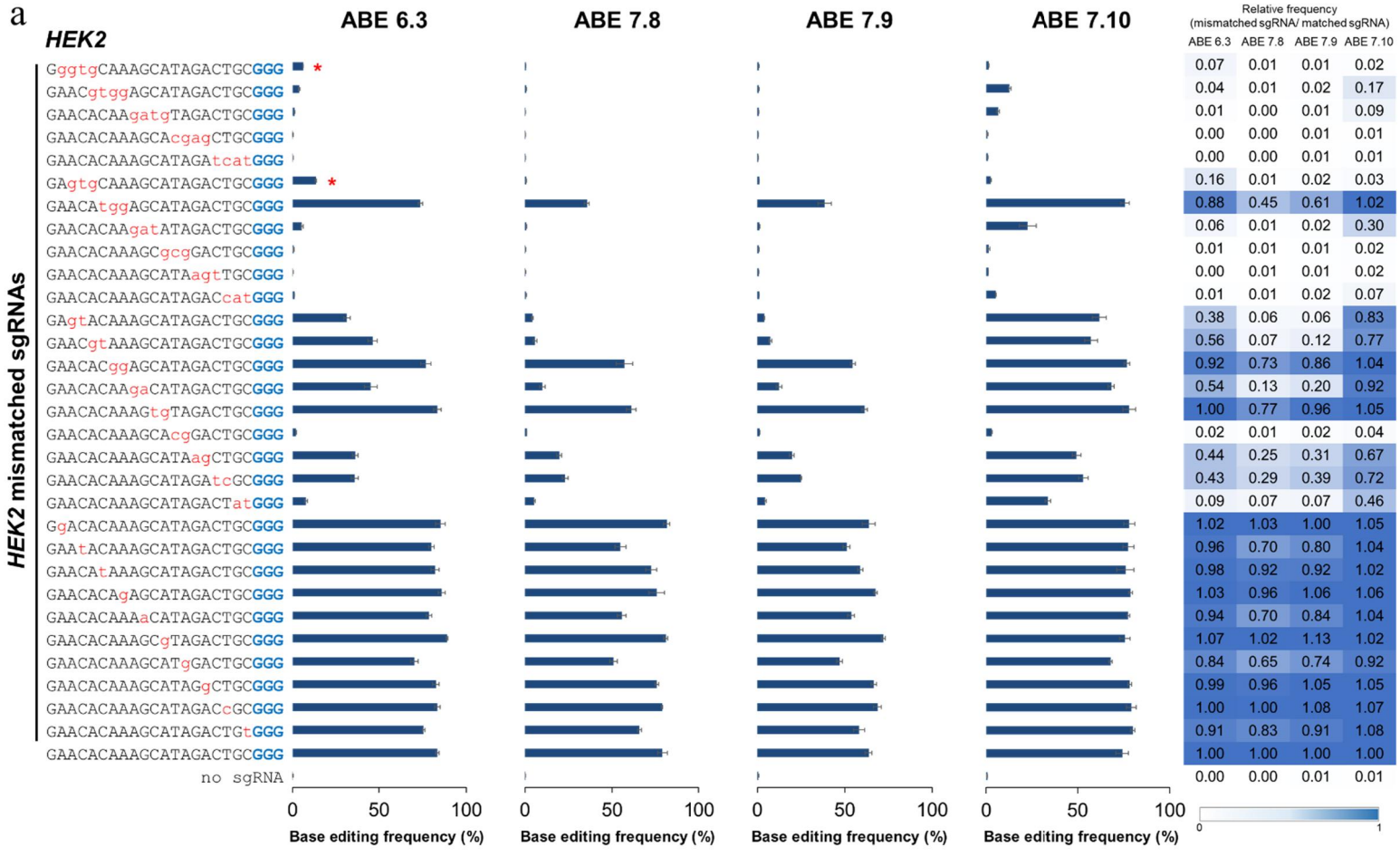

HEK2
R²=0.46
ABE-mediated relative substitution frequency with mismatched sgRNAs
Cas9-mediated relative indel frequency with mismatched sgRNAs
R²=0.88
ABE-mediated relative substitution frequency With mismatched sgRNAs
BE3-mediated relative substitution frequency with mismatched sgRNAs

Figure 4. Correlation between ABE-mediated substitution frequency and Cas9-mediated indel frequency with mismatched sgRNAs for *HEK3* site(a), *RNF2* (b), and *HEK2* site (c). The blue dots indicate that the mismatched sgRNAs with which the ABE-mediated relative base editing frequencies are three times more efficient than Cas9-mediated relative indel frequencies or BE3-mediated relative base editing frequencies. The red color dots indicate that the mismatched sgRNAs with which the Cas9-mediated relative indel frequencies or BE3-mediated relative base editing frequencies are three times more efficient than ABE-mediated relative frequencies.

b. Mismatch tolerance of ABE according to the versions of TadA deaminase mutants

To identify off-target activities of other versions of ABEs, I transfected plasmids encoding sgRNA with one to four nucleotides mismatches and plasmids encoding ABE6.3, ABE7.8, ABE7.9, or ABE7.10 into HEK293T cell lines. In general, the base editing activity of ABE7.10 is known to be highest (Gaudelli et al., 2017), so the base editing frequency was the highest with ABE7.10 using most mismatched sgRNAs, but there were some exceptions (Figure 5a). For example, when using sgRNAs with mismatches at positions 2-5 or 3-5, the activity of ABE6.3 was the highest (indicated with asterisks in Figure 5a). When checking the correlation-efficient value of each pair, I confirmed that ABE7.8 and ABE7.9 were very similar to each other, but relatively less similar to ABE7.10 in *HEK2* site (Figure 5b).


a
HEK2
HEK2 mismatched sgRNAs
ABE 6.3
ABE 7.8
ABE 7.9
ABE 7.10
GggtgCAAAGCATAGACTGCGGG
GAACgtggAGCATAGACTGCGGG
GAACACAAgatgTAGACTGCGGG
GAACACAAAGCAcgagCTGCGGG
GAACACAAAGCATAGAtcatGGG
GAgtgCAAAGCATAGACTGCGGG
GAACAtggAGCATAGACTGCGGG
GAACACAAgatATAGACTGCGGG
GAACACAAAGCgcgGACTGCGGG
GAACACAAAGCATAagtTGCGGG
GAACACAAAGCATAGACcatGGG
GAgtACAAAGCATAGACTGCGGG
GAACgtAAAGCATAGACTGCGGG
GAACACggAGCATAGACTGCGGG
GAACACAAgaCATAGACTGCGGG
GAACACAAAGtgTAGACTGCGGG
GAACACAAAGCAcgGACTGCGGG
GAACACAAAGCATAagCTGCGGG
GAACACAAAGCATAGAtcGCGGG
GAACACAAAGCATAGACTatGGG
GgACACAAAGCATAGACTGCGGG
GAAtACAAAGCATAGACTGCGGG
GAACAtAAAGCATAGACTGCGGG
GAACACAgAGCATAGACTGCGGG
GAACACAAAaCATAGACTGCGGG
GAACACAAAGCgTAGACTGCGGG
GAACACAAAGCATgGACTGCGGG
GAACACAAAGCATAGgCTGCGGG
GAACACAAAGCATAGACcGCGGG
GAACACAAAGCATAGACTGtGGG
GAACACAAAGCATAGACTGCGGG
no sgRNA
*
*
0
50
100
Base editing frequency (%)
Relative frequency
(mismatched sgRNA/ matched sgRNA)
ABE 6.3 ABE 7.8 ABE 7.9 ABE 7.10
0.07 0.01 0.01 0.02
0.04 0.01 0.02 0.17
0.01 0.00 0.01 0.09
0.00 0.00 0.01 0.01
0.00 0.00 0.01 0.01
0.16 0.01 0.02 0.03
0.88 0.45 0.61 1.02
0.06 0.01 0.02 0.30
0.01 0.01 0.01 0.02
0.00 0.01 0.01 0.02
0.01 0.01 0.02 0.07
0.38 0.06 0.06 0.83
0.56 0.07 0.12 0.77
0.92 0.73 0.86 1.04
0.54 0.13 0.20 0.92
1.00 0.77 0.96 1.05
0.02 0.01 0.02 0.04
0.44 0.25 0.31 0.67
0.43 0.29 0.39 0.72
0.09 0.07 0.07 0.46
1.02 1.03 1.00 1.05
0.96 0.70 0.80 1.04
0.98 0.92 0.92 1.02
1.03 0.96 1.06 1.06
0.94 0.70 0.84 1.04
1.07 1.02 1.13 1.02
0.84 0.65 0.74 0.92
0.99 0.96 1.05 1.05
1.00 1.00 1.08 1.07
0.91 0.83 0.91 1.08
1.00 1.00 1.00 1.00
0.00 0.00 0.01 0.01
0
1

b

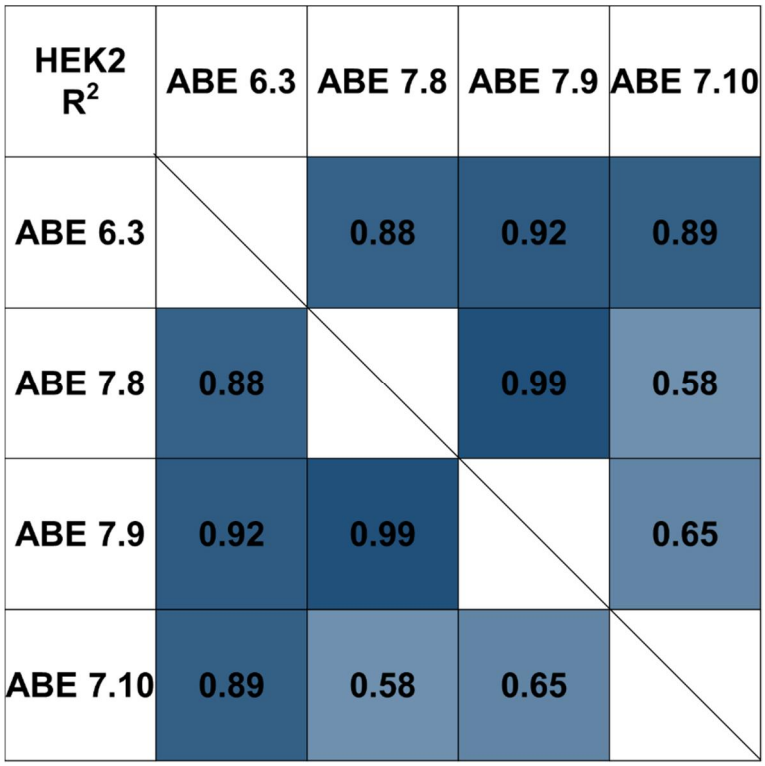


Figure 5. Tolerance of ABE variants-induced base editing frequencies for mismatched sgRNAs. (a) Mismatched sgRNAs that differed from the *HEK2* site by one to four nucleotides were tested in HEK293T cells. Base editing or indel frequencies were measured using targeted deep sequencing. Mismatched nucleotides and the PAM sequence are shown in red and blue, respectively. The red asterisks indicate more highly active mismatched sgRNAs with ABE6.3 versus the other ABE variants including ABE7.10. Means ± s.e.m. were from three independent experiments. (b) Correlation coefficients R^2 value between relative frequencies of several ABE variants with mismatched sgRNAs targeting *HEK2* site.

3. Digenome-seq to identify ABE7.10 off-target sites in human the genome

a. Digenome-seq workflow for the identification of genome-wide ABE7.10 off-target sites

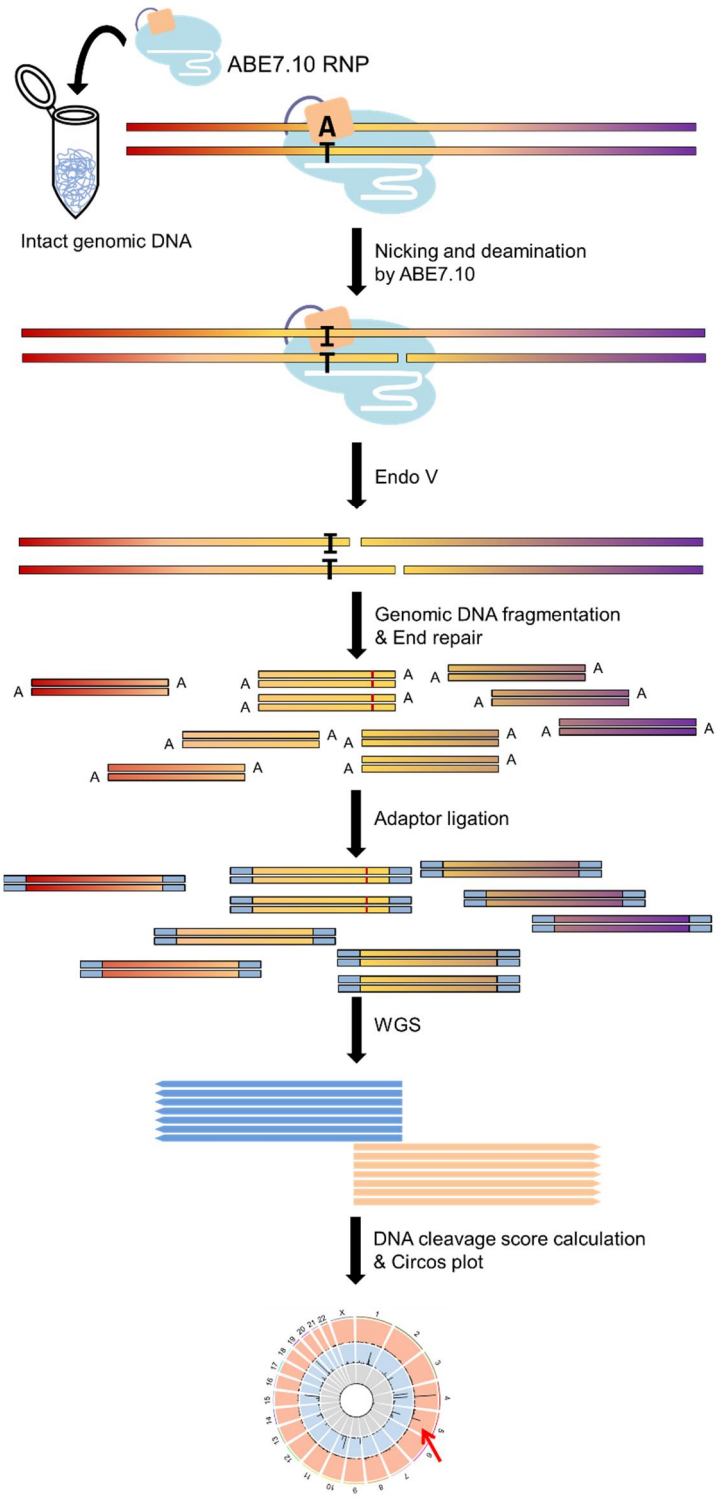
To identify genome-wide off-target sites of ABE7.10 via Digenome-seq, which relies on DNA double-strand breaks (DSBs) produced *in vitro (Kim et al., 2015)*, DSBs should be produced at sites containing inosine, which result from adenine deamination. Note that our group had previously used the recombinant BE3ΔUGI protein (BE3 with no UGI domain) and Uracil-Specific Excision Reagent (USER), a mixture of *E. coli* Uracil DNA glycosylase (UDG) and DNA glycosylase-lyase Endonuclease VIII (Endo VIII), to generate DSBs *in vitro* at sites containing uracil, a deamination product of cytosine, to identify genome-wide BE3 off-target sites comprehensively in an unbiased manner (Kim et al., 2017a). Likewise, *E. coli* Endonuclease V (Endo V) or a combination of human Alkyladenine DNA Glycosylase (hAAG) and Endo VIII could be used to cleave a phosphodiester bond at inosine-containing sites (Figure 6a-c). Endo V recognizes inosine in DNA and catalyzes the cleavage of the second phosphodiester bond 3' to the inosine, whereas hAAG excises hypoxanthine to yield an AP site (apurinic/apyrimidinic site), which is then recognized and processed by Endo VIII to induce a single-strand break. To test this idea, a PCR amplicon containing a target DNA sequence was first treated with the recombinant ABE7.10 protein expressed in and purified from *E. coli* (Figure 7) and its sgRNA to catalyze adenine

deamination in one DNA strand and to produce a nick in the other DNA strand (Figure 6), and then with Endo V (Figure 6b) or with hAAG/Endo VIII (Figure 6c) to yield a composite DSB. As expected, the PCR amplicon was cleaved by ABE7.10 plus Endo V (Figure 8a) or by ABE7.10 plus hAAG/Endo VIII (Figure 9a).

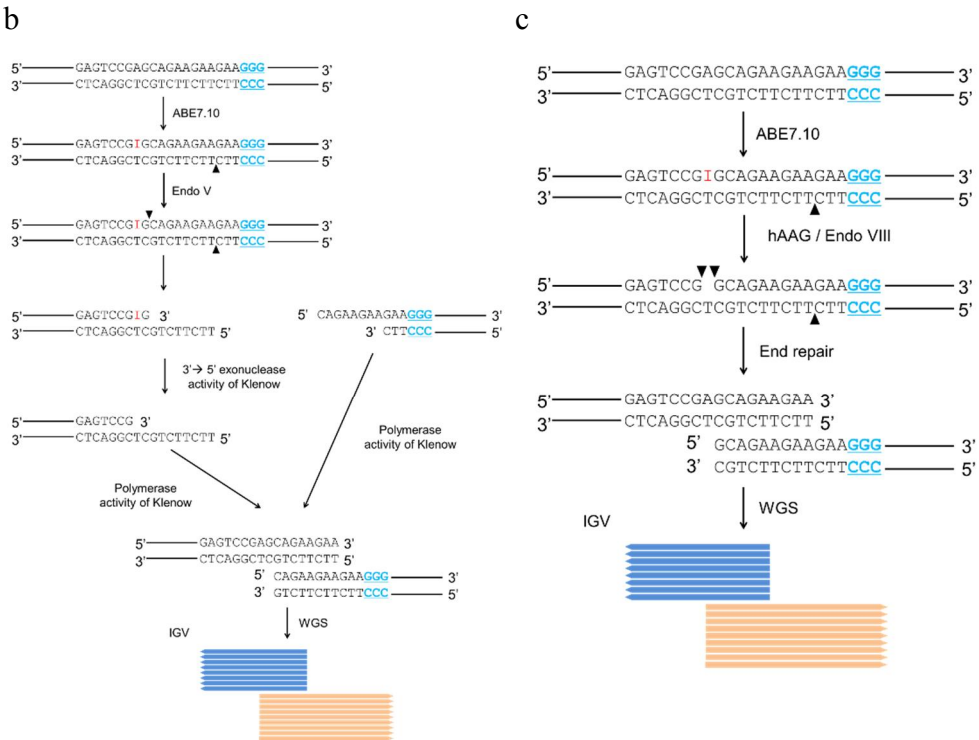
Having confirmed that DSBs could be produced at ABE-targeted sites containing inosine using inosine-specific DNA repair enzymes, we next investigated whether human genomic DNA isolated from HEK293T cells could be processed *in vitro* with ABE7.10 plus Endo V or with ABE7.10 plus hAAG/Endo VIII for Digenome-seq analysis to identify genome-wide ABE off-target sites. Human genomic DNA was incubated with ABE7.10 ribonucleoprotein (RNP) (300 nM ABE7.10 and 900 nM sgRNA) targeting an endogenous chromosomal site (known as *HEK2*) for 8 hours and then with the repair enzymes for 2 hours. Sanger sequencing was used to confirm ABE-mediated adenine-to-inosine conversion (inosine is changed to guanine during PCR amplification) at the target site (the opposite strand with a nick cannot be sequenced) and Endo V-catalyzed DNA cleavage at the inosine-containing site (a nicked DNA strand cannot be amplified) (Figure 8b). When genomic DNA is treated with both ABE7.10 and Endo V, the DNA strand containing inosine is cleaved by Endo V. Note that the DNA strand cleaved by Endo V cannot be amplified by PCR, whereas the DNA strand with no inosine is amplified. As a result, only adenine (and no guanine) is observed at the editing site by Sanger sequencing (Figure 8b). Quantitative real-time PCR were also used to verify targeted DNA cleavage using ABE7.10 plus Endo V (Figure 8c).

Similar results were obtained when hAAG and Endo VIII were used instead of Endo V (Figure 9b-c). The resulting digested genomic DNA samples were subjected to whole genome sequencing (WGS). After mapping sequencing reads to the human reference genome (hg19), alignments of sequence reads were monitored at the on-target site using Integrative Genomics Viewer (IGV) and observed uniform rather than random alignments of sequencing reads, characteristic of a DSB (Figure 8d and Figure 9d). No such alignments were seen with untreated genomic DNA or genomic DNA treated with ABE7.10 alone (Figure 8d and Figure 9d). These results suggest that Digenome-seq can be used for comprehensively mapping genome-wide ABE on-target and off-target sites.

a

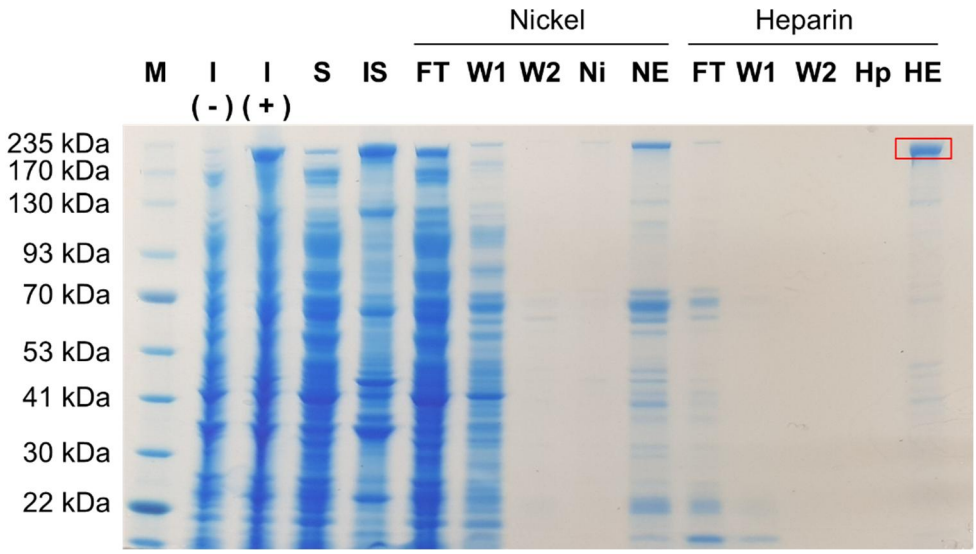

ABE7.10 RNP
A
Intact genomic DNA
Nicking and deamination
by ABE7.10
Endo V
Genomic DNA fragmentation
& End repair
Adaptor ligation
WGS
DNA cleavage score calculation
& Circos plot

(Continued)



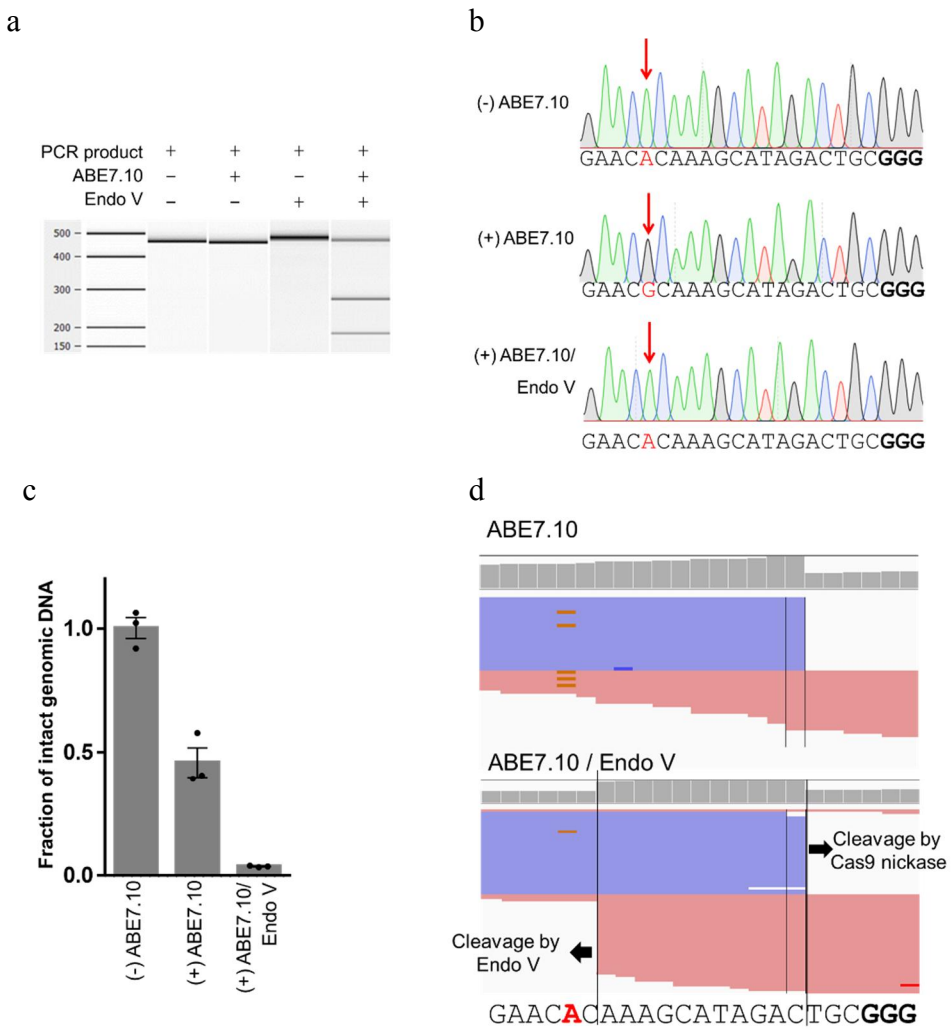
(By Daesik Kim in Institute for Basic Science)

Figure 6. Digenome-seq workflow for the identification of Genome-wide ABE7.10 off-target sites. (a) Overview of the Digenome-seq workflow for ABE-mediated Digenome-seq. (b) Overview of Digenome-seq using ABE7.10. ABE7.10 catalyzes the conversion of guanine to inosine. Endo V recognizes inosine and cleaves the second phosphodiester bond 3' to the inosine. Arrowheads indicate the positions of phosphodiester bonds cleaved by the ABE7.10 nickase and Endo V. (c) ABE7.10 mediates adenine-to-inosine conversion in one strand and produces a nick in the other strand. hAAG excises inosine to produce an AP site and Endo VIII (DNA glycosylase and AP-lyase) cleaves the AP site. Arrowheads indicate the sites cleaved by the ABE7.10 nickase and hAAG/Endo VIII.



(By Gyeorae Lee in University of Science and Technology)

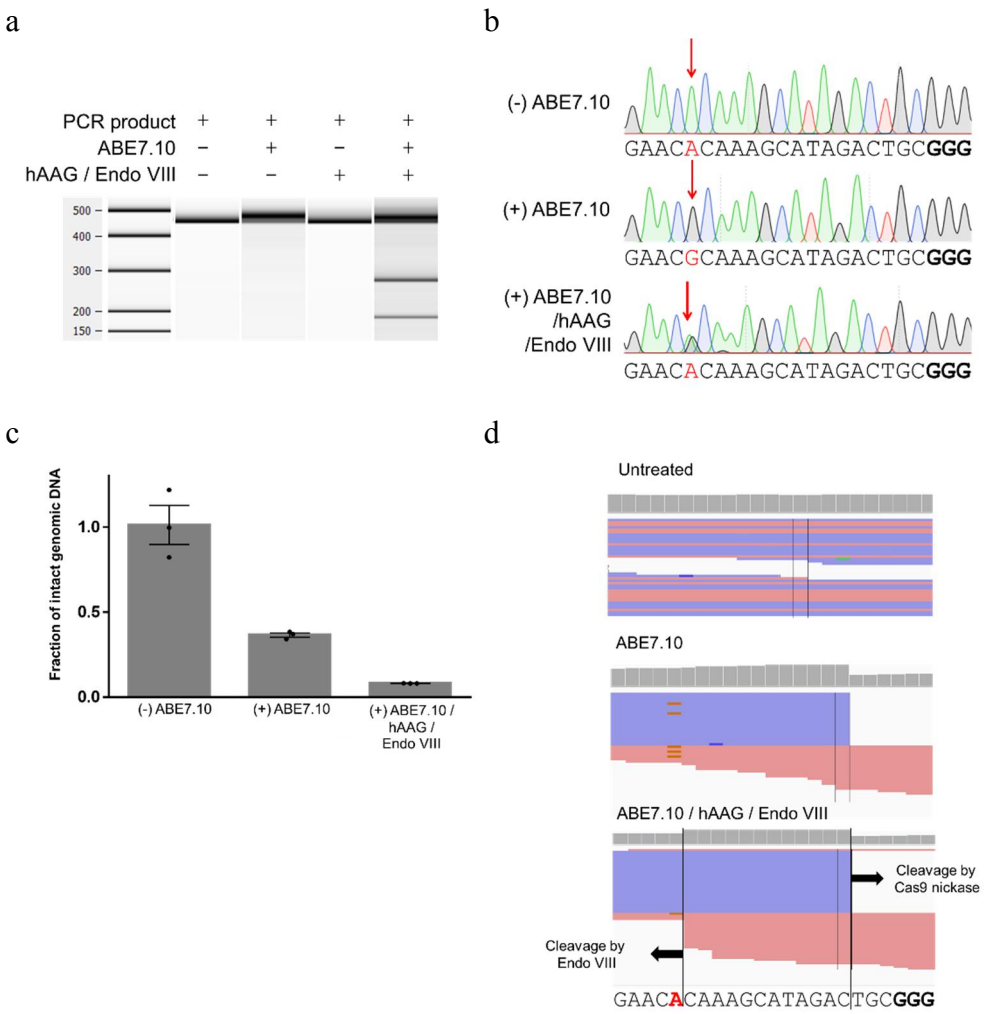
Figure 7. SDS-PAGE analysis of ABE7.10 protein purification using nickel affinity chromatography and heparin bead chromatography. M: marker, I(-): cell lysate before IPTG induction, I(+): cell lysate after IPTG induction, S: soluble lysate fraction, IS: insoluble lysate fraction, FT: flow-through, W1, W2: waste after washing, Ni: Ni-NTA agarose beads after elution of bound protein, NE: protein fraction eluted from nickel beads, Hp: Heparin Sepharose 6 Fast Flow affinity resins after elution of bound protein, HE: protein fraction after purification using heparin beads. The red box represents the ABE7.10 protein band. Similar results were obtained more than three times when the gel electrophoresis was repeated independently.



(By Daesik Kim in Institute for Basic Science and Gyeorae Lee in University of Science and Technology)

Figure 8. ABE7.10-, and Endo V-mediated Digenome-seq. (a) The PCR product containing a ABE7.10 target sequence was cleaved, when treated with both ABE7.10 and Endo V. Similar results were obtained more than three times when the gel electrophoresis was repeated independently. (b) Sanger sequencing results

showing A-to-G conversion by ABE7.10 and DNA cleavage by Endo V. (c) qRT-PCR results showing DNA cleavage by the ABE7.10 nickase and Endo V. Means ± s.e.m. were from three independent experiments. (d) IGV images showing the alignments of sequence reads at the *HEK2* target site obtained using WGS data.



(By Daesik Kim in Institute for Basic Science and

Gyeorae Lee in University of Science and Technology)

Figure 9. ABE7.10, hAAG, and Endo VIII mediated Digenome-seq. (a) A DSB occurs only when the PCR product is treated with ABE7.10, hAAG, and Endo VIII. Similar results were obtained more than three times when the gel electrophoresis was repeated independently. (b) Sanger sequencing results showing an A-to-G conversion when genomic DNA was treated with ABE7.10 and then a G-to-A

conversion when genomic DNA was treated with ABE7.10, hAAG, and Endo VIII. (c) qRT-PCR results showing digestion of genomic DNA by ABE7.10, hAAG, and Endo VIII. Means ± s.e.m. were from three independent experiments. (d) IGV images of the straight alignments of sequencing reads that are observed after treatment with ABE7.10, hAAG, and Endo VIII.

b. Digenome-seq results reveal genome-wide off-target sites of ABE7.10

To determine ABE7.10 off-target sites in the human genome, a DNA cleavage score was assigned to each base-pair position across the entire genome (Figure 10), using the WGS data, and listed all the sites with high scores (the initial cutoff score was set to 2.5 as done previously (Digenome 1.0)), corresponding to *in vitro* cleavage sites (Kim et al., 2016). Note that the Digenome 1.0 program does not consider sequence homology. 17 *in vitro* cleavage sites were obtained using Endo V or 18 such sites using hAAG/Endo VIII (Figure 11a-c). Importantly, 16 sites, including the on-target site, were commonly identified by the two different experimental methods, showing that Digenome-seq is highly reproducible and reliable (Figure 11a-b). Sequence logos, obtained by comparing DNA sequences at these sites, showed that almost every nucleotide position contributed to the specificity of the ABE7.10 deaminase (Figure 11c).

Next, Digenome-seq was performed with each of 6 additional ABE7.10 RNPs. A total of 7 sgRNAs in complex with ABE7.10 plus Endo V cleaved human genomic DNA at one to 28 (8 ± 4 per sgRNA, on average) sites including on-target sites (Figure 12 and Table 4 and 5). Note that our group had shown previously that Cas9 nucleases and BE3ΔUGI deaminases (plus USER) cleave genomic DNA at 70 ± 30 sites (Kim et al., 2016) and 8 ± 3 sites (Kim et al., 2017a), respectively, in the human genome. Among the 7 sgRNAs analyzed in this study, two sgRNAs had been combined with BE3ΔUGI and Cas9 and subjected to Digenome analysis

previously (Kim et al., 2016; Kim et al., 2017a), allowing direct comparisons between adenine and cytosine BEs and Cas9 (Figure 13a-b). The *HEK2* sgRNA yielded 17, 2, and 35 *in vitro* cleavage sites when combined with ABE7.10, BE3, and Cas9, respectively (Figure 13c). The *RNF2* sgRNA yielded 5, 1, and 13 sites with ABE7.10, BE3, and Cas9, respectively (Figure 13d). All of the 3 *in vitro* cleavage sites identified using two BE3 deaminases were also identified using the respective ABE7.10 deaminases. By contrast, *in vitro* cleavage sites obtained with ABE7.10 and Cas9 were largely different. Thus, 10 out of 17 (= 59%) sites obtained with the *HEK2* ABE7.10 were not identified using Cas9. 2 out of 5 (= 40%) sites obtained with the *RNF2* ABE7.10 were not identified using Cas9. These results suggest that ABEs are in general more specific than Cas9 but can recognize different sets of off-target sites.

Then, ABE off-target base editing was validated at these 57 *in vitro* cleavage sites identified above using the 7 sgRNAs via targeted amplicon sequencing. In addition, the Digenome 2.0 program was also used to identify many additional candidate off-target sites with a cutoff score of > 0.1, which contained PAM-like sequences (5'-NGN-3' or 5'-NNG-3') and had 10 or fewer mismatches, compared to the respective on-target sequences. 12 to 152 such sites (60 ± 20, on average) were obtained with the 7 sgRNAs complexed with ABE7.10, which included all of the 7 on-target sites and 57 *in vitro* cleavage sites obtained with Digenome 1.0 (Table 4 and 5).

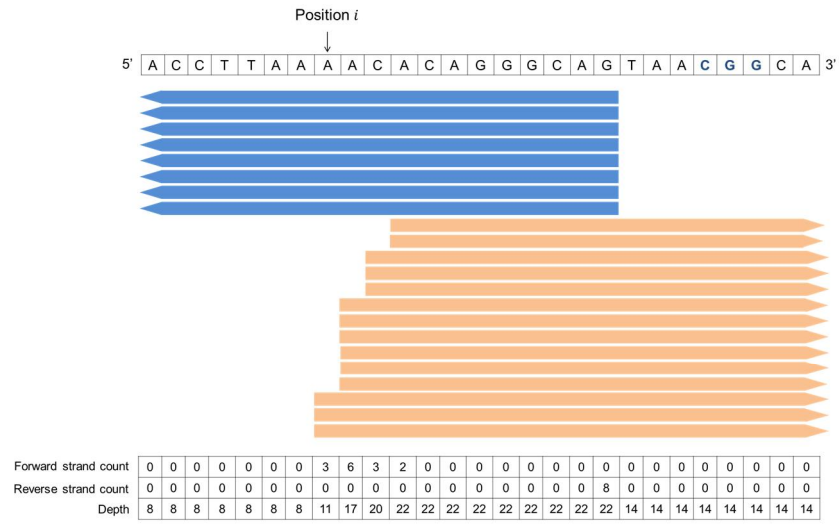
Score at position i =

$$\sum_{a=1}^{5} \frac{(F_i - 1)}{D_i} X \frac{(R_{i+8+a} - 1)}{D_{i+8+a}} \times (F_i + R_{i+8+a} - 2)$$
$$+ \sum_{a=1}^{5} \frac{(R_{i+11} - 1)}{D_{i+11}} X \frac{(F_{i-3+a} - 1)}{D_{i-3+a}} \times (R_{i+11} + F_{i-3+a} - 2)$$

F_i : Number of forward sequence reads starting at position i

R_i : Number of reverse sequence reads starting at position i

D_i : Sequencing depth at position i



Score at position i

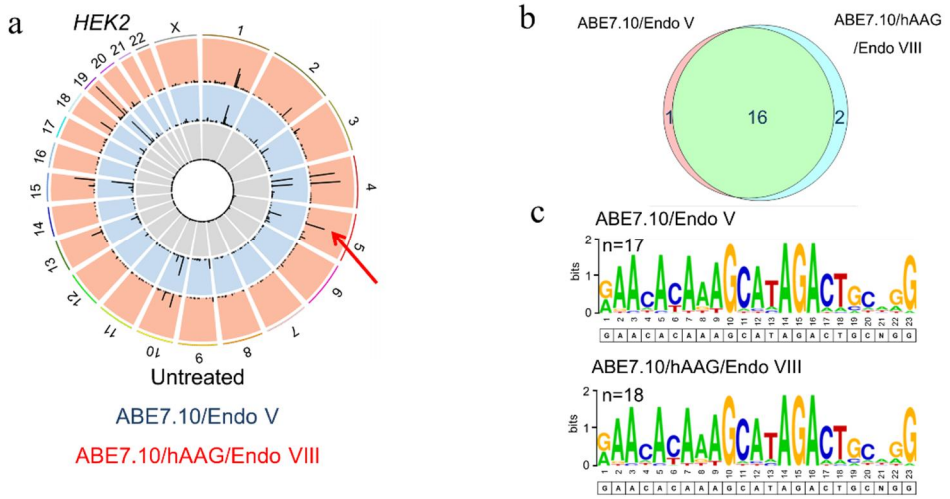
= (6-1)/17 * [(0-1)/22 * (6+0-2) + (0-1)/22 * (6+0-2) + (8-1)/22 * (6+8-2) + (0-1)/14 * (6+0-2) + (0-1)/14 * (6+0-2)]

+ (8-1)/22 * [(0-1)/8 * (8+0-2) + (3-1)/11 * (8+3-2) + (6-1)/17 * (8+6-2) + (3-1)/20 * (8+3-2) + (2-1)/22 * (8+2-2)]

= 2.66

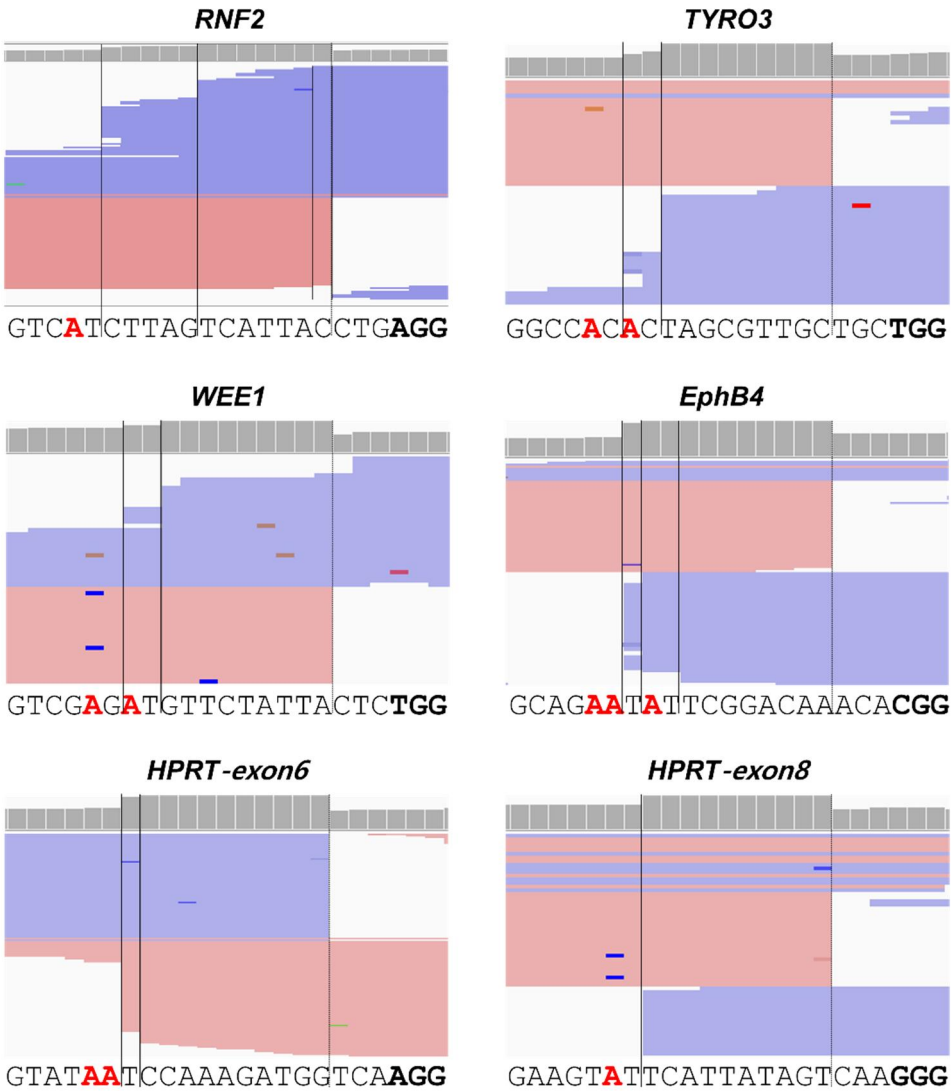
(By Daesik Kim in Institute for Basic Science)

Figure 10. *In vitro* DNA cleavage scoring system for Digenome-seq analysis of ABE.



(By Daesik Kim in Institute for Basic Science)

Figure 11. Genome-wide off-target effects of ABE using ABE7.10- and Endo V- or hAAG and Endo VIII-mediated Digenome-seq (a) Genome-wide Circos plots represent DNA cleavage scores obtained via WGS using intact genomic DNA (gray) and genomic DNA treated with ABE7.10 plus Endo V (blue) or with ABE7.10 plus hAAG and Endo VIII (red). The red arrow indicates the on-target site. (b) A Venn diagram showing the number of *in vitro* cleavage sites in the human genome, identified using ABE7.10 plus Endo V or ABE7.10 plus hAAG and Endo VIII. (c) Sequence logos were obtained using WebLogo by comparing DNA sequences at Digenome-captured sites.



(By Daesik Kim in Institute for Basic Science)

Figure 12. IGV images showing the straight alignments of sequencing reads at the *RNF2*, *TYRO3*, *WEE1*, *EphB4*, *HPRT-exon6*, and *HPRT-exon8* sites.

Table 4. The number of ABE7.10 on-target and off-target sites identified by Digenome-seq

		Number of in vitro cleavage sites	
	On-target DNA sequence	Digenome 1.0	Digenome 2.0
HEK2	GAACACAAAGCATAGACTGCGGG	17	152
RNF2	GTCATCTTAGTCATTACCTGAGG	5	50
TYRO3	GGCCACACTAGCGTTGCTGCTGG	2	12
WEE1	GTCGAGATGTTCTATTACTCTGG	1	21
EphB4	GCAGAATATTCGGACAAACACGG	3	39
HPRT-Exon6	GTATAATCCAAAGATGGTCAAGG	28	148
HPRT-Exon8	GAAGTATTCATTATAGTCAAGGG	1	20
Means ± S.E.M		8 ± 4	60 ± 20

(By Daesik Kim in Institute for Basic Science)

Table 5. Digenome-seq captured sites.

HEK2				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HEK2_001	chr19	35505485	20.1	GAACAC-AAGCAcAGACTGaAGG
HEK2_002	chr15	93557679	15.1	GAACACA-tGCATAGACTGCTAG
HEK2_003	chr4	53536209	14.7	GAAtACtAAGCATAGACTcCAGG
HEK2_004	chr4	90522183	14.3	GAACACAAtGCATAGAtTGCCGG
HEK2_005	chr1	167742859	13.4	aAACACAgAGCAcAGACTGCTGA
HEK2_006	chr10	43159340	13.4	aAACACAAAGaCATAGACcaCTGG
HEK2_007	chr5	87240613	8.9	GAACACAAAGCATAGACTGCGGG
HEK2_008	chr9	290167	6.6	aAACAtAAAGaATAGACTGCAAG
HEK2_009	chr3	28560566	5.8	GAACACAAAGtATAGAaTGCTAG
HEK2_010	chr1	175310138	5.1	GgACACAAAGCtTAGACTcCAGG
HEK2_011	chr13	87616351	4.9	GAAgAtAAAGCATAGACTctAGG
HEK2_012	chr10	87125365	4.6	aAACAtAAAGCATAGACTGCaAAG
HEK2_013	chr18	32279782	3.9	aAAtACAAAGCATAGACTaatATG
HEK2_014	chr2	192248363	3.5	GAACACAtA-CATAGACaGCTGG
HEK2_015	chr5	160202385	3.0	GAAaACAAAGCAaAGAaaGCAGG
HEK2_016	chr15	63236612	2.9	aAcacaAtAGCATAGACTGgaCTG
HEK2_017	chr20	23101377	2.7	atACACAgAGCAaAGACTGCAGG
HEK2_018	chrX	48317375	2.5	aAACAtAAAGCcTAGACTGaCGG
HEK2_019	chr19	49972515	2.2	aAAaACAAAGCgTAGACTGtGGG
HEK2_020	chr12	81785997	2.1	GgACAaAAAGCATAGACTcaAGG
HEK2_021	chr1	40461424	2.1	atACACACAAGCAcAGACTGCAGG
HEK2_022	chr15	92968604	2.1	GAACACAgcaCATAaACTGCAGG
HEK2_023	chr5	142142325	2.0	aAACACtAA-CATAGACTGCAGG
HEK2_024	chr5	131174461	1.7	aAAtACAAtGCATAGACTGCTAG
HEK2_025	chr1	36097072	1.4	GtAaACAAAGCATAGACTGaGGG
HEK2_026	chrX	132724931	1.3	GAACA-tAAtCAcAGACTGCTGG
HEK2_027	chr6	139353017	1.3	ccAaACAAAaCATAGACTGCTGG
HEK2_028	chrX	36949816	1.3	GAAaACAAAaCATAGAgTGCTGG
HEK2_029	chr17	2670110	1.2	aAtCAaAtAGCATAGACTGCATG
HEK2_030	chr13	91254877	1.2	GAACACAAAcatagtACTGaAGG
HEK2_031	chrX	138877515	1.2	aAACAgAAAGCATgGACTGCGGA
HEK2_032	chr13	113428466	1.1	cAAtACAAAGgATAGACTGCAGG
HEK2_033	chr15	41796089	1.1	agACACAcAGCAcAGACTGCAGG
HEK2_034	chr11	128508576	1.0	GAAttCAAAGCATAGAtTGCAGG
HEK2_035	chr11	131870604	0.9	acACACAAAGCATAGACTatGTG
HEK2_036	chr4	91844238	0.9	GgAaACAAAGCATAGACattTGG
HEK2_037	chr3	105801924	0.8	tAAtACAAAGCATAGAtaGtTGG
HEK2_038	chr8	66280643	0.8	GAAtttAAAGCATAGACTGCAAG
HEK2_039	chr3	168165117	0.8	ctgaACAcAGCAaAGACTGCTGG
HEK2_040	chr15	65377019	0.8	GAgCgatAAGCAcAGACTGCTGG

(Continued)

HEK2_041	chr9	97332608	0.7	GtAattAAAGCAcAGACTGCTGG
HEK2_042	chr3	33744669	0.7	GAAaACAAAcCAcAGACTGgGGG
HEK2_043	chr2	19474362	0.7	ctAttgcAAGCAcAGACTGCTGG
HEK2_044	chr16	64198277	0.7	GAAtACAtAaCATAGACTGgGGG
HEK2_045	chr1	199217096	0.7	atACcatAAGCATAGACTGtTGG
HEK2_046	chr22	28895717	0.7	attaAgAtAGCATAGACTGCAGG
HEK2_047	chr13	94827431	0.6	aAtgACtAAtaATAGACTGCTGG
HEK2_048	chr20	895768	0.6	atttcaAcAGCATAGACTGtAGG
HEK2_049	chr1	108869934	0.5	agtatagcAGCATAGACTGCAGG
HEK2_050	chr20	97640	0.5	GAAttCAAAGCATAGAtTGCAGG
HEK2_051	chrX	58256414	0.5	GcAttCtAAtaATAGACTGCTGG
HEK2_052	chr3	10835885	0.5	aAACACAtAGCAcAGcCTGCAGG
HEK2_053	chr8	97317606	0.5	GAACACAgtaCATAGACTGgCAG
HEK2_054	chr7	1973215	0.5	GgctgCctAGCAcAGACTGCCGG
HEK2_055	chr5	145405225	0.5	tctCctAccGCATAGACTGCTGG
HEK2_056	chr14	106066714	0.4	GtcgcaAAcGCATAGACTtCCGG
HEK2_057	chr2	34403831	0.4	cttattAAAGCAcAGACTGCTGG
HEK2_058	chr7	83764326	0.4	actatatAAGCATAGACTGtTGG
HEK2_059	chr14	106188442	0.4	GtcgcaAAcGCATAGACTtCCGG
HEK2_060	chr9	111564240	0.4	cgAtttcAAGCATAGACTGCTGG
HEK2_061	chr4	91393910	0.4	tAAtcatAAGCATAGAaTGCTGG
HEK2_062	chr9	117268271	0.4	GAAtAC-AAGCATAGACTGCTGT
HEK2_063	chr1	208174511	0.4	tcACAaAtAaGCATAGACTGgCGG
HEK2_064	chr6	164073518	0.4	cttgctAAAGCAcAGACTGCTGG
HEK2_065	chr18	56307002	0.4	aAgaACAAAaCATAGACTGCAGG
HEK2_066	chrX	119188148	0.4	aAttAaAcAGaATAGACTGCTGG
HEK2_067	chr10	11472312	0.4	GAACACAgAaCcATAGACTGgGGG
HEK2_068	chrX	93938357	0.4	aAACAaAAAGCATAGACcaCAAG
HEK2_069	chr3	33958845	0.4	GAACACA-gGCATAGACTatGGA
HEK2_070	chr9	90257919	0.4	ctcagagAAGCAcAGACTGCAGG
HEK2_071	chr2	144687377	0.4	atcatatAtGCATAaACTGCAGG
HEK2_072	chr1	191029339	0.4	cAACAaAAtatATAGACTGCTGG
HEK2_073	chr3	124241469	0.3	attgttAAAGCAcAGACTGCAGG
HEK2_074	chr11	80540770	0.3	agAtcttAAGCATAGACTGtGGG
HEK2_075	chr4	146455378	0.3	tAACACAAtacCATAGACTGgCGG
HEK2_076	chr6	70806985	0.3	GAAtACAAAtaATAGACTatTGG
HEK2_077	chr3	10539384	0.3	GAAtttAAAGCATAGACTctGGG
HEK2_078	chr13	80619390	0.3	aAACAaAtAaaATAGACTGCAGG
HEK2_079	chr5	126385455	0.3	ccACACcAAGCATAGACTtCTGG
HEK2_080	chr9	92683115	0.3	aAcaAaAcAGCATAGACTGCATG
HEK2_081	chr17	2810242	0.3	tAAttgcAAGCATAGACgGCAGG
HEK2_082	chr12	131431445	0.3	GgAgAgAgAGCATAGACTGCTGG
HEK2_083	chr11	105148271	0.3	agAaAtAAtatATAGACTGCAGG

(Continued)

HEK2_084	chr1	75043747	0.3	tggttactAGCATAGACTtCAGG
HEK2_085	chr20	31946581	0.3	aAACACttAaCAcAGACTGCAGG
HEK2_086	chr8	58172690	0.3	ccACcagcAGCATAGACTcCAGG
HEK2_087	chr7	39877158	0.2	aAgacattttCATAGACTGtCGG
HEK2_088	chr3	65508026	0.2	tAAgcacAAGCATAGgCTGCTGG
HEK2_089	chr6	40925009	0.2	aAcaAgtAtGCATAGACTGCTGG
HEK2_090	chr15	81265977	0.2	aAAtACcAAcCAcAGACTGCAGG
HEK2_091	chr8	102768749	0.2	aAAtgCAtAGCATAGACTGCTGA
HEK2_092	chr1	237916998	0.2	tAACACAAAGacTAaACTGCAGG
HEK2_093	chr1	215323272	0.2	tAcCACgAAGCATAGACTGtAGG
HEK2_094	chr14	60249634	0.2	GAtCACtcAaCATAGACTGCATG
HEK2_095	chr3	29416510	0.2	GAACAg-AAaCATAGACcaaTGG
HEK2_096	chr11	79574188	0.2	tAcaAaAtAagATAGACTGCTGG
HEK2_097	chr10	96064388	0.2	acAgttAAAGaATAGACTGCAGG
HEK2_098	chr7	78443524	0.2	aAACACAAAgGCATAaACTcCAGG
HEK2_099	chr13	51093353	0.2	GAACACAcA-CAcAGACTGaGGG
HEK2_100	chr13	45266353	0.2	GgACAtAAAtCATAaACTGCTGG
HEK2_101	chr1	20011125	0.2	aAAgctAcAGCATAGACTGCAAG
HEK2_102	chr1	77190607	0.2	tcACACAAAcCATAGACTGaGGG
HEK2_103	chr11	68639975	0.2	GAtCtacctatATAGACTGCAGG
HEK2_104	chr9	31652144	0.2	tAtagaggAGCATAGACTGCTGG
HEK2_105	chr19	5068807	0.2	tAtCAgAAAGCAcAGACTGCGGG
HEK2_106	chr12	54405073	0.2	GAACAacAAGgATAaACTGCCGG
HEK2_107	chr12	33389548	0.2	tAcaAacAcGCATAGAaTGCCGG
HEK2_108	chr1	15519301	0.2	tAAgAaAtgGaATAGACTGCAGG
HEK2_109	chr20	21064111	0.2	aAAtAttAAGCATAGACTaCGGG
HEK2_110	chr20	17801623	0.2	GAAtgCcAAGCATAaACTGCAGG
HEK2_111	chr2	19844956	0.2	aActcCAAAGCATAtACTGCTGG
HEK2_112	chr2	116312126	0.2	GtgaACAcAGCATAGACTGgGGC
HEK2_113	chr15	35048979	0.2	atcCtaAgAaCATAGACTGCAGG
HEK2_114	chr11	19349648	0.2	tttgttAAAaCATAGACTGCTGG
HEK2_115	chr4	170242082	0.2	tAACAaAAtGCATAGACTGCTAG
HEK2_116	chr1	66454108	0.2	tgtataAcAGCATAcACTGCTGG
HEK2_117	chr9	20221246	0.2	aAACACtAAtaATAGACTGtGGG
HEK2_118	chr12	2617156	0.2	aAgCAatcAGCAcAGACTGCGGG
HEK2_119	chr19	32225924	0.2	GgACACAgAGtATAGACTGaGGG
HEK2_120	chr6	123554091	0.1	cAcaAatcAaaATAGACTGCTGG
HEK2_121	chr13	58440671	0.1	GAAttCAAAGCATgGACTGCAGG
HEK2_122	chr19	18370525	0.1	ttcCAaAAAaCATAGACTGCTGG
HEK2_123	chr8	74980814	0.1	GtAgAgtAAGCATAcACTGCTGG
HEK2_124	chrX	101465228	0.1	GgcCtttAgGCAcAGACTGCAGG
HEK2_125	chr21	16057037	0.1	aAgactttAGCATAGACTGCAAG
HEK2_126	chr5	76655640	0.1	GAAaAgtAAG-ATAGACTGCCGG

(Continued)

HEK2_127	chrX	37275566	0.1	aActcatcAGCATAGACTGtAGG
HEK2_128	chrX	39813705	0.1	aAAtACAgAGCtTAGACTaCTGG
HEK2_129	chr3	99649762	0.1	aAACAtgAAGCATAGACaGCAAG
HEK2_130	chr3	86016290	0.1	acAtAgAtAttATAGACTGCTGG
HEK2_131	chr2	42080736	0.1	cAcattAAAGCAcAGACTGCTGG
HEK2_132	chr7	154994106	0.1	GAgCctAgAGCAcAGACTGCAGG
HEK2_133	chr10	85566693	0.1	aAgacttAAGCAaAGACTGCTGG
HEK2_134	chr1	176037520	0.1	ttACACAtAGCAcAGACTaCAGG
HEK2_135	chr7	56616286	0.1	agAagttAAGCATAGACTGgTGG
HEK2_136	chrX	69481649	0.1	ctAtAttgAGaATAGACTGCAGG
HEK2_137	chr17	79546106	0.1	GtgCACAcAGCATAGACTGCATG
HEK2_138	chr9	74099753	0.1	ctAgcaAAAaCATAGACTGCTGG
HEK2_139	chr7	113235709	0.1	tAAaACAAAGaAcAGACTGCTGG
HEK2_140	chr5	165353994	0.1	tAcagatAgtaATAGACTGCAGG
HEK2_141	chr5	134255399	0.1	aAttAatAAcaATAGACTGCTGG
HEK2_142	chr6	39115051	0.1	aAgattgcAGCATAaACTGCTGG
HEK2_143	chr10	732884	0.1	aAgCACAtAaCATAGACTGaAGG
HEK2_144	chr14	88566211	0.1	GtAgAaAAAGtATAGACTGCAGG
HEK2_145	chr14	89446986	0.1	tActAtggtGCATAGACTGCTGG
HEK2_146	chr11	133057022	0.1	ctctACtgAGaATAGACTGCAGG
HEK2_147	chr13	47277564	0.1	tgAaAtAAtaCATAaACTGCCGG
HEK2_148	chr6	42655828	0.1	aAtaAacttGCATAGACTGtAGG
HEK2_149	chr10	72922827	0.1	agACcCtAAGCATAGACTGCAGA
HEK2_150	chr16	55456878	0.1	tAcCcaAAtGCAcAGACTGCTGG
HEK2_151	chr2	194416373	0.1	aAgatagAtttATAGACTGCTGG
HEK2_152	chr4	177589824	0.1	aAACACAAAaCATAaAtTaCTGG

RNF2				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
RNF2-01	chr1	210858376	7.3	tcaccaTTAGTCATTACCTGCTG
RNF2-02	chr1	185056773	4.4	GTCATCTTAGTCATTACCTGAGG
RNF2-03	chr17	53928598	4.3	GTCATCTTAGTCATTAC-TGAGG
RNF2-04	chr6	143212078	4.1	GTaATaTTAGTCATTACCgGTGG
RNF2-05	chr17	17072075	3.6	GTCATCcTAGTCATTtaCTGGGG
RNF2-06	chr12	131619804	2.2	aTCAcCTTAGcCATTACCaGGGG
RNF2-07	chr9	102127746	1.4	GTCA-CTTAGTCATTgCCTGTGG
RNF2-08	chr15	92315580	1.0	GTCAcaTTAGcCATTACCTGTGA
RNF2-09	chr2	57559709	1.0	tgtATCTTAGaCATTACaTGTGG
RNF2-10	chr5	29673802	0.8	aTttTCTTAGTgATTAtCTGGGG
RNF2-11	chr15	25135467	0.8	catcTaaTAGTaATTACCTGGGG
RNF2-12	chr3	60339996	0.6	aTagTaTTAGTCATTACCTGTGA
RNF2-13	chr12	12595622	0.4	acCATCTTAGTCATTAtCTaATG
RNF2-14	chrX	131090978	0.4	aTttTCTTAGTCATTACCTaGAG

(Continued)

RNF2-15	chr2	51894204	0.4	aTCATCTTcaTCATTACaTGAGG
RNF2-16	chr2	134638146	0.4	cctATaTataTCATTACCTtTGG
RNF2-17	chr16	62531206	0.4	GTCATtTTAGTCATTAtCTtGAG
RNF2-18	chr22	44137982	0.4	aTCAcCTgAGTCATTACCcaTGG
RNF2-19	chr4	125521739	0.4	tTaATCTTAGTCATTACtTTtTGG
RNF2-20	chr2	102668931	0.3	aTCATCaTcGTCATTAtCTGGGG
RNF2-21	chr14	88336132	0.3	aTCA--TTAGTCATTgCCTGAGG
RNF2-22	chr7	43967266	0.3	cTtAT-TTAGTCATTACCTGTAG
RNF2-23	chr2	41520178	0.3	GTCATaTTAaTCATTACaTaGAG
RNF2-24	chr1	111785389	0.3	tcCATCTcAcTCATTACCTGAGG
RNF2-25	chr8	51846780	0.3	GTtATCTTAGTCtTTACCTGAGA
RNF2-26	chr2	24749046	0.3	caCATCTTAccCATTACaTGAGG
RNF2-27	chr6	5869239	0.3	GcCAcCTcAGTCATTAgCTGGGG
RNF2-28	chrX	133138307	0.3	acatTagTAaaCATTACCTaGGG
RNF2-29	chr1	59227209	0.3	taCcTacctGTCATTACCTaTGG
RNF2-30	chr2	177556597	0.3	GatATCTTAGcCATTACCTaGGA
RNF2-31	chr3	113984053	0.3	tatcTacTAcTCATTACCTGAGG
RNF2-32	chr2	66353253	0.2	agCATCTctGTCATTACCcaGGG
RNF2-33	chr8	99451745	0.2	cgtgcaTTAGTCATTACCTGAGG
RNF2-34	chr5	109304296	0.2	GTCtgtaTAGTCATTACCTtTGG
RNF2-35	chr6	155177402	0.2	taCATCTatGTtATTACCTaTGG
RNF2-36	chr17	58390142	0.2	tTCATCTTtGTCATTACCTaAAG
RNF2-37	chr11	8498488	0.2	tatActaTAtaCATTACCTGGGG
RNF2-38	chr6	3726562	0.2	GctcatTcAcTCATTACCTaTGG
RNF2-39	chr16	5977420	0.2	GTgATCTaaAGTCATTACCTtAGG
RNF2-40	chrX	76946717	0.2	tTCAataTAaTCATTACCTGTGG
RNF2-41	chr18	62061226	0.2	aTtATtTTAGTCATTACCTtTGG
RNF2-42	chr19	14867049	0.2	ccCATCTcAGcCATTACCTGGGT
RNF2-43	chr13	70399620	0.1	agaATaTTAGTCcTTACCTGGGG
RNF2-44	chr12	86219647	0.1	GatcctTTAaTCATTACCTtTGG
RNF2-45	chr13	103624473	0.1	acCATCTTAGTCAcTACCTGGGC
RNF2-46	chr1	174478567	0.1	GTatTaTTgtaCATTACCTGAGG
RNF2-47	chr2	209871803	0.1	agCATCTTAGcCATTACCTcTAG
RNF2-48	chr5	91144278	0.1	agCATCTTAaTCcTTACCTcAGG
RNF2-49	chr4	160106364	0.1	GaatcCTcAtaCATTACCTGTGG
RNF2-50	chr4	37552932	0.1	aTCATCTTttTaATTACCTaTGG

TYRO3				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
TYRO3-01	chr15	76553332	4.3	GGCCACACTAGtGTTGCcGCTGG
TYRO3-02	chr15	41857272	3.4	GGCCACACTAGCGTTGCTGCTGG
TYRO3-03	chr20	45312757	2.0	GGCCACACcAGCcTTGCTGtCGG
TYRO3-04	chr5	173133468	0.8	GGCCAC--TAGCGTTGCTcCAGG

(Continued)

TYRO3-05	chr22	50686658	0.6	GGCCACACTgaCcTTGCTGCTGG
TYRO3-06	chr9	140109027	0.4	ttCCACAC-AGCtTTGCTGCTGG
TYRO3-07	chr17	17207899	0.3	GGCCAC-CTAGgGTTGCTGCTGG
TYRO3-08	chr11	99666699	0.2	ccaCcCtCTAGCaTTGCTGCTGG
TYRO3-09	chr11	27495183	0.2	cctaACACcAaCGTTGCTGCTGG
TYRO3-10	chr1	38684664	0.2	ctgagCtCTAGCaTTGCTGCTGG
TYRO3-11	chr3	22936710	0.1	GGCatataTAGCaTTGCTGCTGG
TYRO3-12	chr6	150629124	0.1	GaaCACACTAaCaTTGCTGtGGG

WEE1				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
WEE1-01	chr1	234241713	11.1	tTaGAGATGTTCTATTAtTCCGG
WEE1-02	chr11	9610086	2.0	GTCGAGATGTTCTATTACTCTGG
WEE1-03	chr4	18855558	1.9	GagGAGATGcTCTATTACTCCGG
WEE1-04	chrX	42756785	0.7	aTCaAGATGaaCTATTACTCTGG
WEE1-05	chr1	215805743	0.4	GTgGAGATGTTaTgTTACTCTGG
WEE1-06	chr5	57863350	0.4	GaCataATGTTCTATTACTCAAG
WEE1-07	chr4	119010382	0.3	taatAGATGTTCTATTACTaAGG
WEE1-08	chr15	80845387	0.3	aTtaAGATGTTCTATTAaTaGGG
WEE1-09	chr13	87867165	0.2	cagctaATGTaaTATTACTCAGG
WEE1-10	chr4	107037600	0.2	aTCaAGATaTTCTATTACTgGGG
WEE1-11	chrX	136297269	0.2	tcttAtATtTTCTATTACTCAGG
WEE1-12	chr1	8500986	0.2	aTaGAGcaGTTaTATTACTCTGG
WEE1-13	chr11	71806684	0.2	GTaGAGATGTTCTcTaACTCAGG
WEE1-14	chr12	83599167	0.2	cTCttttatTTCTATTACTCTGG
WEE1-15	chr8	71394697	0.1	aatataATaTTCTATTACTCAGG
WEE1-16	chr8	86942319	0.1	caactataaTTCTATTACTCTGG
WEE1-17	chr1	61165563	0.1	GgCtActgtcTCTATTACTCAGG
WEE1-18	chr20	45730192	0.1	aTttAGATaTTCTATTACTCTGA
WEE1-19	chr8	134795474	0.1	cTttAagataTCTATTACTCTGG
WEE1-20	chr2	128238175	0.1	caaGAGAcGcTCTATcACTCAGG
WEE1-21	chr7	31677975	0.1	aaacAGcTGTTCTATTACTCAGG

EphB4				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
EphB4-01	chr7	100411300	10.3	GCAGAATATTCGGACAAACACGG
EphB4-02	chr5	81021421	8.0	GCAGAATATTaCaGACAAACtAGG
EphB4-03	chr15	47712791	4.1	GCAGAATATcaGGACAAACAATG
EphB4-04	chr11	123329727	2.4	aCAGAATATTCaGACAAtCACAG
EphB4-05	chr1	93047514	2.2	GCAGtA-AaTaaGACAAACAAGG
EphB4-06	chr15	51198151	1.9	aCAaAATATTCGGACAggCACGG
EphB4-07	chr18	29712895	1.8	GCAGAATtTaaGGACAAgCAAGG
EphB4-08	chr3	55670295	1.8	GCAGAATA--aGGACAAACATGG

(Continued)

EphB4-09	chr13	25919264	1.4	cagtAATATTCaGACtAAACATGG
EphB4-10	chr3	122057177	0.9	ttAatATATaaGGACAAACATGG
EphB4-11	chr1	26891812	0.8	aCAtAATATTCaaACAAACAGGG
EphB4-12	chr10	131074805	0.6	aatGAATAaagGGACAAACAGGG
EphB4-13	chr4	123425007	0.5	tCAG-ATATaaGGACAAACATGG
EphB4-14	chr5	60692862	0.4	GCtGAATATTgGGACAAtCATGG
EphB4-15	chr9	930844	0.4	GCAGttATAaaaaGACAAACAGGG
EphB4-16	chr9	29344386	0.4	aCAGAATATaaGaACAAAtATGG
EphB4-17	chr4	160752620	0.4	atAGAATAaTCaGACAAAaAAGG
EphB4-18	chr16	23706770	0.4	GCAG-ATAaagGGACAAACATGG
EphB4-19	chr11	117354211	0.3	GacagtgAaTaGGACAAACATGG
EphB4-20	chr14	90497703	0.3	tgAGAATATgCGGACAAACACAG
EphB4-21	chr2	68459706	0.3	atAGAATATTaGGACAAAtAGAG
EphB4-22	chr13	51951355	0.2	caAtAATAaaaGGACAAACATGG
EphB4-23	chr17	7777056	0.2	aCAGgATAaaaGGACAAAtATGG
EphB4-24	chr3	188431448	0.2	aCAGAtATAcgtaGACAAACAGAG
EphB4-25	chr12	11705824	0.2	atAGAATAagaaGACAAACATGG
EphB4-26	chr10	95697506	0.2	ttAGAATATTCctACAAACAAGG
EphB4-27	chrX	123665071	0.2	cCAtgtTATTCaGACAAACATGG
EphB4-28	chr16	75869847	0.2	GCAGAA-AaTaGGACAAtCACGG
EphB4-29	chr6	128784409	0.2	agAGAATATa-GGACAAACtTGG
EphB4-30	chr5	144297821	0.2	GCAGAAataTttGACAAACAAGG
EphB4-31	chr14	29760524	0.1	aaAGAATATa-GaACAAACAGGG
EphB4-32	chr1	36914336	0.1	aCAGAA-ATaCGaACAAACATGG
EphB4-33	chr21	26227853	0.1	aacaAATgcaaGGACAAACATGG
EphB4-34	chr13	63528393	0.1	aCActAT-TTaaGACAAACAGGG
EphB4-35	chr10	72516747	0.1	aCAGAActacgaGACAAACACGG
EphB4-36	chr7	20108292	0.1	GCAGA-TAccCaGACAAACAGGG
EphB4-37	chr1	89896856	0.1	taAGAATAT-aGGACAAACAGGA
EphB4-38	chr4	145533104	0.1	cCAGAATATTa-GACAAACATGG
EphB4-39	chr20	9784096	0.1	agAGAATATagGGACAAAtAAGG

HPRT-Exon6				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HPRT_E6-001	chrX	133627607	29.0	GTATAATCCAAAGATGGTCAAGG
HPRT_E6-002	chr12	7468915	21.3	GTATAcatCcAAGATGGcCAGGG
HPRT_E6-003	chr3	10057621	17.0	GTATtA-CCAAAGATGGTCtGGG
HPRT_E6-004	chr22	39586062	12.7	GTATAATC-AAAGATGGcCcTGG
HPRT_E6-005	chr7	13928792	10.5	GTATtAT-CAAAaATGGTCAAGG
HPRT_E6-006	chr1	225831201	10.2	aTATAA-CCAAAGATGtTCACAG
HPRT_E6-007	chr17	53575925	9.3	aTATAATCttAAGATGGTCAAGG
HPRT_E6-008	chr12	13406973	9.0	aTATAATCCtAAGATGtTCATGG
HPRT_E6-009	chr14	27255077	6.1	GTATtATCCAAAGATGacCAGAG

(Continued)

HPRT_E6-010	chr7	85914581	5.6	aTAcAATCCAAAGtTGGTCtGGG
HPRT_E6-011	chr3	9906633	4.7	aTATAttaCCAAAGATGGTCtGGG
HPRT_E6-012	chr8	125496169	4.5	GTAgAATCCAtAGATGGaCAGGG
HPRT_E6-013	chr17	10064482	4.3	GTATtA-CCAAAGATGGTCATGG
HPRT_E6-014	chr2	53374352	4.3	GTAcAATCaAAAGATGaaCAAGG
HPRT_E6-015	chr4	102026801	4.1	aTATtATCCAAAGATGGagtTGG
HPRT_E6-016	chr6	19070067	3.8	aTgTAtTCCAAAGATGGTCAGGG
HPRT_E6-017	chr6	7761387	3.7	aagTAATCCAAAGATaGTCtTGG
HPRT_E6-018	chr12	59613203	3.6	GTATAAcCCAAAGAaaGTaATGG
HPRT_E6-019	chr16	20336811	3.6	aTATAtATtCcAAGATGGTtcTGG
HPRT_E6-020	chr4	136809212	3.5	tTATAATCCAAAGAaGGcCAAGG
HPRT_E6-021	chr12	117915656	3.5	GTATctAcCCAAAGATGtTCATGG
HPRT_E6-022	chr14	45005861	3.4	GTATAAT--AAAGATGGTCAAGG
HPRT_E6-023	chr14	104033102	3.3	aTAgAtTCCAAAGATGGaCATGG
HPRT_E6-024	chr11	93731924	3.1	cTATAATCtAAAaATGGTCAAGG
HPRT_E6-025	chr13	75257165	2.9	GTATAtTagAAAGATGGTCATGA
HPRT_E6-026	chr15	54911620	2.8	GgATAATCaAAAGATGGaCtTGG
HPRT_E6-027	chrX	6849511	2.6	aTATAATgCAAAGcTGGTCACGG
HPRT_E6-028	chr11	87526371	2.5	aTAcAATCCAAAGATatTCtAGG
HPRT_E6-029	chr3	19652583	2.4	GTATAATCCAAAGgTGGcCtAGG
HPRT_E6-030	chr17	45009128	2.4	aTATAATCCAtAGAaGGTCAAGA
HPRT_E6-031	chr8	9090663	2.3	GatatAatacAAGATGGaCAAGG
HPRT_E6-032	chr12	6810501	2.2	GgATAATCCAAAGATGGTgcAAG
HPRT_E6-033	chr6	118233456	1.5	aTATAAaCCAAAGATcGTtAAGG
HPRT_E6-034	chr14	85667729	1.4	tctcccTCCAAAGAaGGTCACGG
HPRT_E6-035	chrX	110551547	1.4	tTATAgTCCAAAGgaaGTCAGGG
HPRT_E6-036	chr8	18630339	1.2	tatTtATCCAAAGATGGTCtAGA
HPRT_E6-037	chr12	45338334	1.1	aTATAtaCCAAAGgTaGTCATGG
HPRT_E6-038	chr7	135056752	1.1	aggagATCCAAAGATGGTCAAGG
HPRT_E6-039	chr9	82090375	1.1	acATAgTCCAAAtATGGaCAGGG
HPRT_E6-040	chr12	103187212	1.0	GTATAtAattcAAGATGGTCAGGA
HPRT_E6-041	chr12	58996606	1.0	GaATAAgCCAAAGATGGTCAGTG
HPRT_E6-042	chr12	49657468	1.0	GTATctcCCAgAGATGGTCAAGG
HPRT_E6-043	chr7	120024149	0.9	aTgTAATCtAAAGATGaTtATGG
HPRT_E6-044	chr15	101375955	0.9	GTATAA-CCAAtGATGGTCAGGA
HPRT_E6-045	chr10	13873222	0.9	aTgTcATCCAAAGAaGGTCAGAG
HPRT_E6-046	chr21	41428166	0.9	agATAtTCCAAAGATGGTgAGGG
HPRT_E6-047	chr1	7314483	0.8	tTtccAcCCAAAGATGGTCcAGG
HPRT_E6-048	chr8	133182863	0.8	tTATAgctCcAAGATGGTCAAGG
HPRT_E6-049	chr6	1253997	0.8	tcAgAATCCtAAGATGGTCAGGA
HPRT_E6-050	chr12	15783424	0.8	GgATAATCCAAAGtTGGTCATAG
HPRT_E6-051	chr9	15160111	0.7	GTATAATCC-tAGATGGTgAGGG
HPRT_E6-052	chr12	29153372	0.7	GTATttTCCAAcGATGGTCATGG

(Continued)

HPRT_E6-053	chr3	1123204	0.7	GaATAATtCAAAGATaaTCAAGG
HPRT_E6-054	chr2	50999543	0.7	aaATAATa-AAAGATGGTCATGG
HPRT_E6-055	chr2	36768513	0.7	cTcatATCCAAAGATGGTaAAGG
HPRT_E6-056	chr6	4210565	0.7	tTATAA-CCAAAGATGGTaATGG
HPRT_E6-057	chr14	48260581	0.7	GccTAATCCAAAGATGagCAGGG
HPRT_E6-058	chr8	117575569	0.7	cTATAATgCCAAAGATGGTCATGG
HPRT_E6-059	chr5	160988739	0.7	GTATAgatCAAAGATGccCACGG
HPRT_E6-060	chr14	94364216	0.6	tctcctTCCAAAGATGGTCtGGG
HPRT_E6-061	chr9	125328946	0.6	aTAatcTCCAAAGATGtTCAGGG
HPRT_E6-062	chr9	134083092	0.6	aTATttTCCAAAGATGaTCACAG
HPRT_E6-063	chr16	6159271	0.6	tatTgATCCAAAGATGaTCAAGA
HPRT_E6-064	chr22	27822987	0.6	aTATAtTtCcAAGATGGgCAGGG
HPRT_E6-065	chr19	48640167	0.6	tTcTcATCCAAgGATGGTCAGGG
HPRT_E6-066	chr11	107913428	0.6	aTATAAaCaAAAGATGGTttAGG
HPRT_E6-067	chr22	25519897	0.6	aaATAgTCCAAgGATGGcCAGGG
HPRT_E6-068	chr2	222954701	0.5	GTATttTCtAAAGATGGTCATGA
HPRT_E6-069	chr1	21736973	0.5	GgATAAatCcAAGtTGGTCATGG
HPRT_E6-070	chr20	41218527	0.5	aTATttTCCAAAGATGGTCATGA
HPRT_E6-071	chr16	55855323	0.5	tgActcTCCAAAGATGGTCACAG
HPRT_E6-072	chr3	39725483	0.5	caATAtTCCAAAGATGGTtATGG
HPRT_E6-073	chrX	141390834	0.5	GcAgcATCCAAAGATGGgCAGGG
HPRT_E6-074	chr4	23070418	0.5	cTATAATtCAAAGATGtTCtTGG
HPRT_E6-075	chr8	9653862	0.5	cacatATatAAAGATGGTCATGG
HPRT_E6-076	chr22	32085052	0.5	GTATctTCCAAAGATGGcCAATG
HPRT_E6-077	chr16	8011143	0.5	aatattaCCtAAGATGGTCATGG
HPRT_E6-078	chr13	21574593	0.5	GTATAAaCCAAAaATatTCAGGG
HPRT_E6-079	chr12	58995581	0.4	GaATAAgCCAAAGATGGTCAGTG
HPRT_E6-080	chr10	107263858	0.4	aTATAATgCAAAGAaaGTCAAAG
HPRT_E6-081	chr11	44024764	0.4	GTATAAgCCAAAGcaGGTCACGG
HPRT_E6-082	chr6	138575959	0.4	aTATAAaCCAAAGAcGGcCcTGG
HPRT_E6-083	chr3	98960373	0.4	agATAATCCAAgGATGGTagAGG
HPRT_E6-084	chr8	43181791	0.4	aTATAATCCAgAaATGtTCACGG
HPRT_E6-085	chr5	118103678	0.4	actcttTtCcAAGATGGTCAGGG
HPRT_E6-086	chr7	129222414	0.4	cTgatgTCCAAAGATGGTCtCGG
HPRT_E6-087	chr3	160736673	0.4	GTgaAtTCCAAAGAaGGTCATGG
HPRT_E6-088	chrX	125678217	0.3	catTAATCCAAAGATGaTaAAGG
HPRT_E6-089	chr6	152915585	0.3	GctctATCCAAAGATGGaCAGGG
HPRT_E6-090	chr9	128839841	0.3	cTgTAATCCAgAaATaGTCAGGG
HPRT_E6-091	chr14	96980434	0.3	GTATAtgTCCAAAGATGaTCATGC
HPRT_E6-092	chr13	80168480	0.3	cTATAATCCAAgGAaGGcCAAGG
HPRT_E6-093	chr8	74179696	0.3	GTATccTCCAAAGATGagtAAGG
HPRT_E6-094	chr4	112559644	0.3	aaATAATaCAAAGATGGTtgGGG
HPRT_E6-095	chr13	74505576	0.3	caATAATCCAAAGAaGaTtAAGG

(Continued)

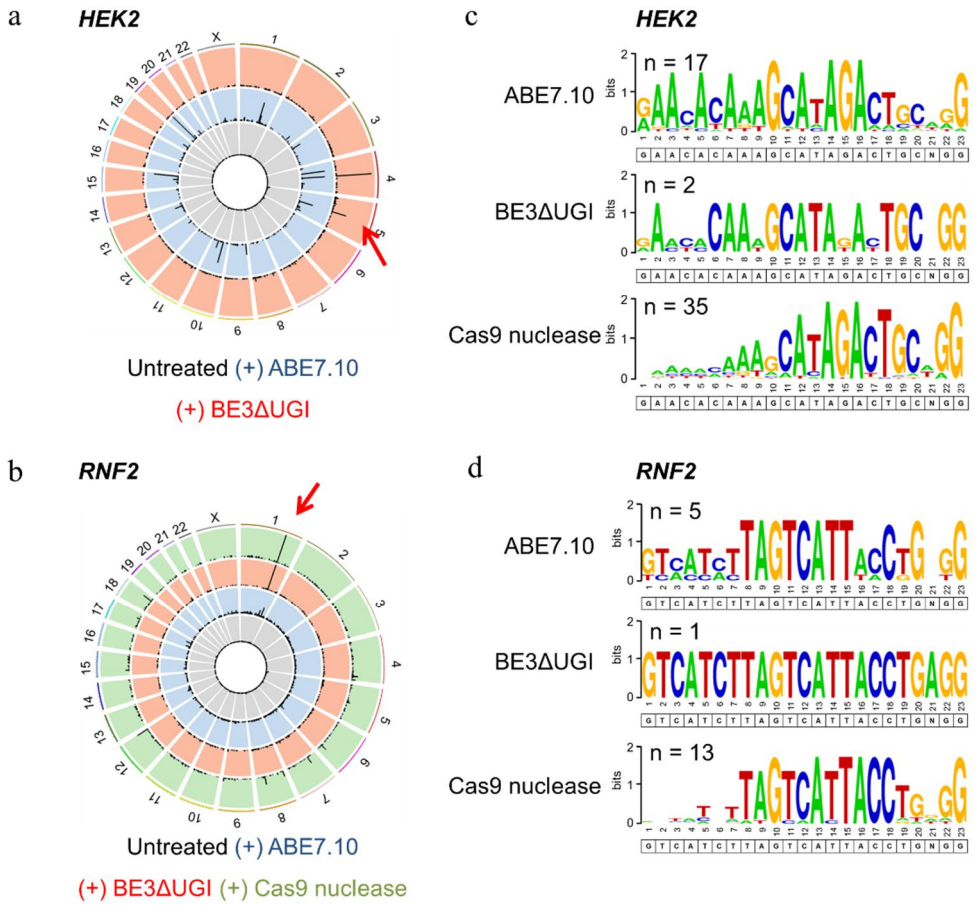
HPRT_E6-096	chr19	32267421	0.3	aTATAATCaCAAAGATGGcaATGG
HPRT_E6-097	chr14	95877075	0.3	GTtTAAgCCAAAGtTGGTCgGGG
HPRT_E6-098	chr8	36263307	0.3	GTAgcATCCAAAGATGGTggAGG
HPRT_E6-099	chr5	28287911	0.3	GTAgAATatcAAGATGGTCAAGG
HPRT_E6-100	chr5	98310114	0.3	agAatgcCCtAAGATGGTCtGGG
HPRT_E6-101	chr6	76199451	0.3	aTATAAcCagAAGATGtTCATGG
HPRT_E6-102	chr10	83966021	0.3	GTATtATCtAAAGgaGGTCATGG
HPRT_E6-103	chr1	55051337	0.2	agATAtaCagAAGATGGTCgGGG
HPRT_E6-104	chr8	91475188	0.2	aaATAATCCAAAaATaGTCATGG
HPRT_E6-105	chr9	20225753	0.2	cTtTAtTCCAAAGATaGTCATGG
HPRT_E6-106	chr6	81303564	0.2	GcATAttTCCAAAcATGGTCAAGG
HPRT_E6-107	chr1	241707903	0.2	GTATttTtCAAAGATGGTCATGA
HPRT_E6-108	chr9	28194167	0.2	GTATttTCCAAAGATGGcCACGA
HPRT_E6-109	chr3	17000484	0.2	tataAATCCAAgGATGGTCtGAG
HPRT_E6-110	chr16	73853834	0.2	tTATttTCCAAAGATGGTCACGA
HPRT_E6-111	chr15	29999660	0.2	tgcacATCCAcAGATGGTCAGGG
HPRT_E6-112	chr5	157715478	0.2	GTgTAATCCAgAGATGaTCACAG
HPRT_E6-113	chr10	2711644	0.2	tctTAATCCAtAGAaGGTCAAGG
HPRT_E6-114	chr17	61689903	0.2	tTcaAAT--AAAGATGGTCACGG
HPRT_E6-115	chrX	125134402	0.2	tcATAATCCAAAaGATGGaCAGGG
HPRT_E6-116	chr1	94251248	0.2	aTATAATaCAAAataGGTCAGGG
HPRT_E6-117	chr22	35100488	0.2	cTAgAtaCCAAAGATGGTCAGAG
HPRT_E6-118	chr2	38407889	0.2	aTtTAgTCCAAAGAgGGTCAGAG
HPRT_E6-119	chr7	112015061	0.2	aTATAcAaCCAAAGATGtTCtAGG
HPRT_E6-120	chr20	2163127	0.2	GTATAtAcatAAAGAaGGTCATGG
HPRT_E6-121	chr8	100230333	0.2	aTtatATCCAAAtATGGTCAGAG
HPRT_E6-122	chr2	206041685	0.2	tTATAATCCAtAAaATGGTCATAG
HPRT_E6-123	chr1	153531237	0.2	tTAgtcTCCAAAGATGGTCACTG
HPRT_E6-124	chrX	118562021	0.2	tatcttTCCAAAGATGGTCtATG
HPRT_E6-125	chr10	26718853	0.2	GTATttTCCAAAGcTaGTCATGG
HPRT_E6-126	chr17	61180425	0.2	GcATAATCCAAgGAaGGTCcTGG
HPRT_E6-127	chr2	125810813	0.2	aTATttTCCAAAaATGGTCACAG
HPRT_E6-128	chr7	28299854	0.2	aTATAATCCAAAagaaGgaAAAG
HPRT_E6-129	chr11	23995596	0.1	aTATAtcCaAAAGATGGTCtGGA
HPRT_E6-130	chr1	34307709	0.1	GatattaaCAAAGATGGTCtGGG
HPRT_E6-131	chr5	173751304	0.1	GaAgAATCCAAcGATGGcCACGG
HPRT_E6-132	chrX	14442705	0.1	acATAATCaCAAAtATGGTCACGG
HPRT_E6-133	chr20	15867186	0.1	GTATttTCCAAAGATGGcCACAG
HPRT_E6-134	chr5	157484648	0.1	tctaAtgaatAAGATGGTCgAGG
HPRT_E6-135	chr1	203269364	0.1	agATAATgCAAAGAcGGTtAGGG
HPRT_E6-136	chr12	28980303	0.1	GTATAtaCCAAAaATGGTgAAGG
HPRT_E6-137	chr18	52818481	0.1	aTATAc-CCAAAGATGGTCcAAG
HPRT_E6-138	chr10	131217430	0.1	tctgAcTCCcAAGATGGTCtTGG

(Continued)

HPRT_E6-139	chr3	109018997	0.1	GgAgAtaCCAAAGATGGTCAGGA
HPRT_E6-140	chr21	20933404	0.1	GTATAtTtCAAAtATGGTCATAG
HPRT_E6-141	chr2	220647283	0.1	aTATAA-CCAAAGATtGTCAGAG
HPRT_E6-142	chr22	23199165	0.1	cTggAgTCCAgAGATGGTCATGG
HPRT_E6-143	chr6	17237684	0.1	GTAgAAaCCAgAGATGGaCAAGG
HPRT_E6-144	chr10	82243028	0.1	tTATAATaCAAAGgTGGTaATGG
HPRT_E6-145	chr15	43004167	0.1	cTATAATCaAAAGtaGGTCATAG
HPRT_E6-146	chr7	25485727	0.1	aTATttTCCAAAaATGGTCATAG
HPRT_E6-147	chr2	71429668	0.1	GTAaAATCCAAAGAgGacCAGGG
HPRT_E6-148	chr3	141311087	0.1	cctaAATCtAAAGATGGTCgGGG

HPRT-Exon8				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites
HPRT_E8-01	chrX	133632686	2.6	GAAGTATTCATTATAGTCAAGGG
HPRT_E8-02	chr11	93382674	1.3	GAAGcATTCATTATAGTCAAAGG
HPRT_E8-03	chr21	34111536	0.6	GAAGaATTCATTATAGaCAATGG
HPRT_E8-04	chr11	100880874	0.6	aAgcTATTCATTATAGcaAATGG
HPRT_E8-05	chr22	29285654	0.6	GAAGTtaATTCATTATAGaCAATGG
HPRT_E8-06	chr4	15866630	0.5	aAAGTATTaATTATAGTCAAGGA
HPRT_E8-07	chr2	53543615	0.5	GAAGTAT-CcTTATAGTCAgAAG
HPRT_E8-08	chr3	149876617	0.4	GAAGTAT-CATTATAGTCttGGG
HPRT_E8-09	chr8	71429233	0.4	GcAGTATTCATaATAGTCAAAAG
HPRT_E8-10	chrX	146032194	0.4	aAAGTATTtcaTATAGcCAAAGG
HPRT_E8-11	chr5	132311070	0.4	GAAtTATTCATcATAGcCAAAGG
HPRT_E8-12	chr3	36349740	0.3	tAAGTAaTCATTATAGTCAgATG
HPRT_E8-13	chr3	181655989	0.3	atAGTATTCATTATAGTaAcAGG
HPRT_E8-14	chr11	116029576	0.3	aAAaTATcCATTATAGTCttACG
HPRT_E8-15	chr10	131905650	0.3	GAAcTATTCATTATtaTCAACGG
HPRT_E8-16	chr9	89312308	0.2	GttGTATTCATTATAGTCAAGAG
HPRT_E8-17	chr4	178543184	0.2	atAGTATTCATTATAGTaAcAGG
HPRT_E8-18	chr12	72733535	0.2	agAGTATcCAccATAGTCAAGGG
HPRT_E8-19	chr1	45916291	0.1	GAAGTAcatcTTATAaTCAATGG
HPRT_E8-20	chr4	153599265	0.1	aAtGTATTCATTATAGgCAAGAG

(By Daesik Kim in Institute for Basic Science)



(By Daesik Kim in Institute for Basic Science)

Figure 13. Genome-wide specificities of ABE7.10, BE3, and Cas9. (a-b) Genome-wide Circos plots for *HEK2* (a) and *RNF2* (b) sites obtained with intact genomic DNA (gray) and genomic DNA treated with ABE7.10/Endo V (blue), BE3ΔUGI/USER (red), or Cas9 (Green). Arrows indicate the two on-target sites. (c-d) Sequence logos were obtained using WebLogo by comparing DNA sequences at Digenome-captured sites.

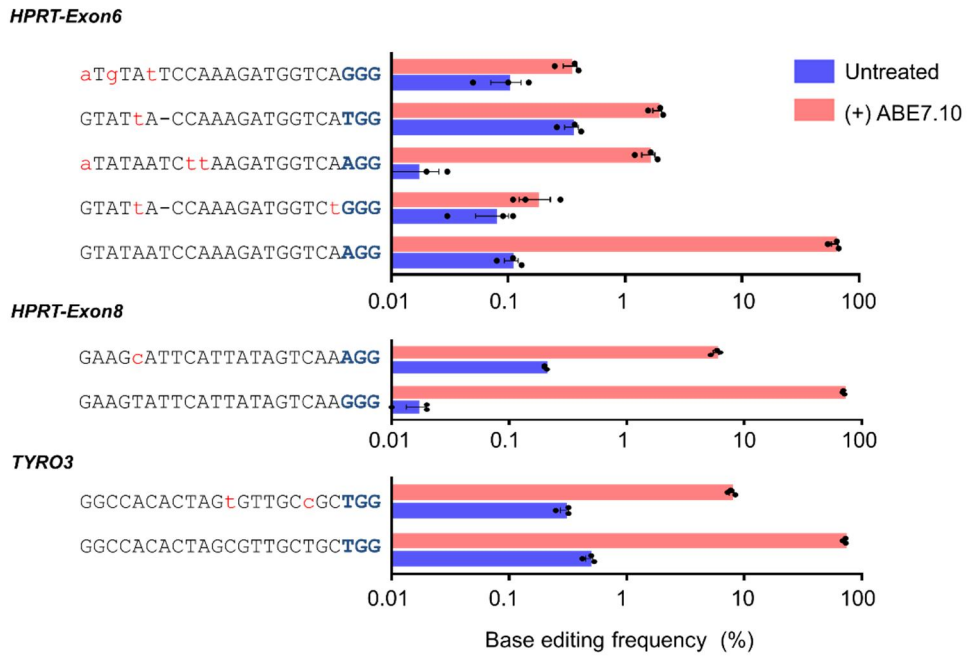
4. Validation of genome-wide ABE7.10 off-target effects

The base editing frequencies of the 7 ABE7.10 deaminases were measured at a total of 193 sites, including those with high sequence homology or with a cleavage score of at least 1.0, via deep sequencing (Figure 14a and Table 6). Single-nucleotide substitutions were not measurably detected with frequencies above noise levels (typically in the range of 0.1~1%) caused by sequencing errors at 173 sites. Among the other 20 validated sites, the 7 on-target sites were base-edited at frequencies that ranged from 29% to 72%. The other 13 validated off-target sites were edited much less efficiently, at frequencies that ranged from 0.1% to 7.8%. This result shows that Digenome-seq is sensitive enough to capture ABE off-target sites with very low editing frequencies that are near detection limits.

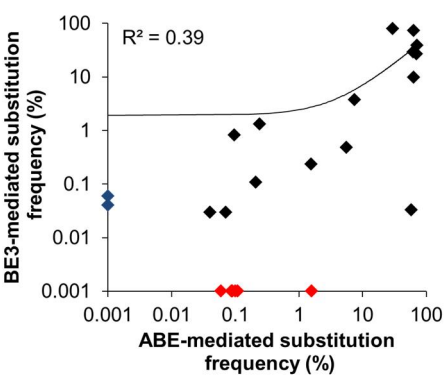
In parallel, Cas9-induced indel frequencies and BE3-induced substitution frequencies were also measured at the 193 sites to compare ABE7.10 with BE3 and Cas9. Unexpectedly, ABE7.10 activities were correlated more strongly with Cas9 activities (R^2 = 0.77) than with BE3 activities (R^2 = 0.39). Among six sites edited at relatively high frequencies with ABE7.10 but poorly with BE3 (Base editing frequencies of 0.001% or lower) (shown as red dots in Figure 14b), two sites had no cytosine within the BE3 editing window of several nucleotides. Seven sites were edited efficiently with ABE7.10 but not with Cas9 (indel frequencies of 0.001% or lower) (shown as red dots in Figure 14c). Among these, three sites (43% = 3/7) had a single extra or missing nucleotide compared to their respective on-target sites, possibly forming a DNA or RNA bulge, respectively, when hybridized with their respective sgRNAs. Our group had also reported previously that such bulge-

forming sites are common among BE3 off-target sites (Kim et al., 2017a) but rare among Cas9 off-target sites (Kim et al., 2016). Among the 193 sites identified via Digenome-seq using ABE7.10, two or four sites were edited in cells at relatively high frequencies with BE3 or Cas9, respectively, but poorly with ABE7.10 (substitution frequencies of 0.001% or lower) (shown as blue dots in Figure 14b, c), showing again that ABE7.10, BE3, and Cas9 can recognize different sets of off-target sites in the human genome.

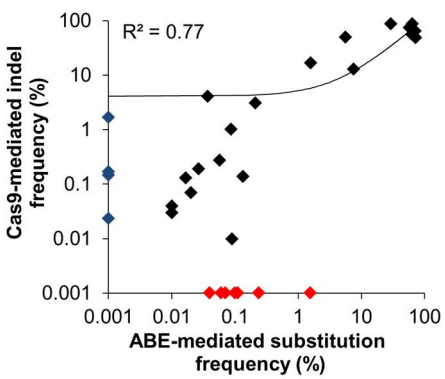
a



b



c



(With Daesik Kim in Institute for Basic Science)

Figure 14. Genome-wide off-target validations of ABE7.10, BE3, and Cas9. (a) Base editing efficiencies at ABE7.10 off-target sites measured using targeted deep sequencing. PAM sequences are shown in blue. Mismatched bases are shown in red.

Dashes indicate RNA bulges. Means ± s.e.m. were from three independent experiments. (b-c) Scatterplots showing correlations between ABE7.10-mediated substitution frequencies and BE3-mediated substitution frequencies (n = 23) (b) or between ABE7.10-mediated substitution frequencies and Cas9-mediated substitution frequencies (n = 32) (c) Pearson's correlation coefficient was used to determine the correlation. The blue and red dots indicate off-targets validated for Cas9 or BE3 but not for ABE7.10 and those validated for ABE7.10 but not for Cas9 or BE3, respectively.

Table 6. Mutation frequencies of ABE7.10, BE3, and Cas9 at on-target and off-target sites captured by Digenome-seq.

						Base editing frequency (%)			Indel frequency (%)	
		Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
HEK2	HEK2_001	chr19	35505485	20.1	GAACAC-AAGCAcAGACTCaAGG	0.12	0.11	0.06	0.04	0.01	0.01
	HEK2_002	chr15	93557679	15.1	GAACACA-tGCATAGACTGCTAG	0.00	0.00	0.04	0.04	0.02	0.01
	HEK2_003	chr4	53536209	14.7	GAAtACtAAGCATAGACTcCAGG	0.06	0.09	0.03	0.04	0.00	0.01
	HEK2_004	chr4	90522183	14.3	GAACACAAtGCATAGAtTGCCGG	0.10	0.31	0.07	0.18	0.02	3.09
	HEK2_005	chr1	167742859	13.4	aAACACAgAGCAcAGACTCCTGA	0.15	0.13	0.00	0.00	0.00	0.01
	HEK2_006	chr10	43159340	13.4	aAACACAAAGaCATAGACcaCTGG	0.00	0.00	0.00	0.00	0.06	0.11
	HEK2_007	chr5	87240613	8.9	GAACACAAAGCATAGACTGCGGG	0.06	63.19	0.06	74.03	0.00	87.57
	HEK2_008	chr9	290167	6.6	aAACAtAAAGaATAGACTGCAAG	0.06	0.12	0.03	0.03	0.02	0.01
	HEK2_009	chr3	28560566	5.8	GAACACAAAGtATAGAaTGCTAG	0.10	0.09	0.00	0.00	0.00	0.00
	HEK2_010	chr1	175310138	5.1	GgACACAAAGCtTAGACTcCAGG	0.00	0.00	0.03	0.03	0.00	0.01
	HEK2_011	chr13	87616351	4.9	GAAgAtAAAGCATAGACTctAGG	0.09	0.08	0.00	0.00	0.00	0.00
	HEK2_012	chr10	87125365	4.6	aAACAtAAAGCATAGACTGCaAAG	0.05	0.06	0.00	0.00	0.02	0.00
	HEK2_013	chr18	32279782	3.9	aAAtACAAAGCATAGACTaatATG	0.00	0.00	0.00	0.00	0.01	0.00
	HEK2_014	chr2	192248363	3.5	GAACACAtA-CATAGACaGCTGG	0.08	0.08	0.18	0.11	0.00	0.00
	HEK2_015	chr5	160202385	3.0	GAAaACAAAGCAcAAGAaaGCAGG	0.11	0.13	0.00	0.00	0.00	0.00
	HEK2_016	chr15	63236612	2.9	aAcacaAtAGCATAGACTGgaCTG	No PCR					
	HEK2_017	chr20	23101377	2.7	atACACAgAGCAaAGACTCCAGG	0.13	0.07	0.00	0.00	0.01	0.00
	HEK2_018	chrX	48317375	2.5	aAACAtAAAGCcTAGACTGaCGG	0.18	0.20	0.08	0.06	0.00	0.00
	HEK2_019	chr19	49972515	2.2	aAAaACAAAGCgTAGACTGtGGG	0.12	0.12	0.00	0.00	0.00	0.00
	HEK2_020	chr12	81785997	2.1	GgACAaAAAGCATAGACTcaAGG	0.12	0.20	0.02	0.03	0.09	0.08
	HEK2_021	chr1	40461424	2.1	atACACACAAGCAcAGACTGCAGG	0.17	0.19	0.23	0.23	0.01	0.08
	HEK2_022	chr15	92968604	2.1	GAACACAgcaCATAaACTGCAGG	0.13	0.11	0.08	0.08	0.00	0.01
	HEK2_023	chr5	142142325	2.0	aAACACtAA-CATAGACTCCAGG	0.09	0.09	0.00	0.00	0.09	0.06
	HEK2_024	chr5	131174461	1.7	aAAtACAAtGCATAGACTGCTAG	0.00	0.00	0.08	0.06	0.00	0.00
	HEK2_025	chr1	36097072	1.4	GtAaACAAAGCATAGACTGaGGG	0.13	0.14	0.03	0.05	0.01	0.01
	HEK2_026	chrX	132724931	1.3	GAACA-tAAtCAcAGACTCCTGG	0.13	0.16	0.08	0.08	0.00	0.00
	HEK2_027	chr6	139353017	1.3	ccAaACAAAaCATAGACTCCTGG	0.15	0.16	0.06	0.03	0.00	0.00
	HEK2_028	chrX	36949816	1.3	GAAaACAAAaCATAGAgTGCTGG	0.00	0.00	0.03	0.02	0.01	0.02
	HEK2_029	chr17	2670110	1.2	aAtCAaAtAGCATAGACTGCATG	0.23	0.15	0.00	0.00	0.00	0.00
	HEK2_030	chr13	91254877	1.2	GAACACAAAcatagtACTGaAGG	0.09	0.11	0.00	0.00	0.01	0.02
	HEK2_031	chrX	138877515	1.2	aAACAgAAAGCATgGACTGCGGA	0.09	0.12	0.02	0.03	0.01	0.00
	HEK2_032	chr13	113428466	1.1	cAAtACAAAGgATAGACTGCAGG	0.05	0.09	0.04	0.06	0.00	0.00
	HEK2_033	chr15	41796089	1.1	agACACAcAGCAcAGACTGCAGG	0.11	0.22	0.07	0.04	0.00	0.00
	HEK2_034	chr11	128508576	1.0	GAAtcCAAAGCATAGAtTGCAGG	0.05	0.05	0.07	0.13	2.42	4.10
	HEK2_042	chr3	33744669	0.7	GAAaACAAAcCAcAGACTGgGGG	0.03	0.04	0.07	0.11	0.03	0.07
	HEK2_044	chr16	64198277	0.7	GAAtACAtAaCATAGACTGgGGG	0.00	0.00	0.08	0.03	0.00	0.00
	HEK2_050	chr20	97640	0.5	GAAtcCAAAGCATAGAtTGCAGG	0.22	0.19	0.00	0.00	0.22	0.23
	HEK2_052	chr3	10835885	0.5	aAACACAtAGCAcAGcCTGCAGG	0.13	0.11	0.03	0.02	0.13	0.11

(Continued)

HEK2_065	chr18	56307002	0.4	aAgaACAAAaCATAGACTGCAGG	0.11	0.15	0.15	0.08	0.11	0.07
HEK2_079	chr5	126385455	0.3	ccACACcAAGCATAGACTtCTGG	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_082	chr12	131431445	0.3	GgAGAgAGCATAGACTGCTGG	0.00	0.00	0.34	0.24	0.00	0.00
HEK2_092	chr1	237916998	0.2	tAACACAAAGacTAaACTGCAGG	0.11	0.08	0.12	0.11	0.11	0.09
HEK2_093	chr1	215323272	0.2	tAcCACgAAGCATAGACTGtAGG	0.02	0.03	0.14	0.13	0.02	0.05
HEK2_099	chr13	51093353	0.2	GAACACACA-CAcAGACTGaGGG	0.06	0.08	0.00	0.00	0.13	0.08
HEK2_101	chr1	20011125	0.2	aAAgctAcAGCATAGACTGCAAG	0.00	0.00	0.07	0.06	0.00	0.00
HEK2_105	chr19	5068807	0.2	tAtCAgAAAGCAcAGACTGCGGG	0.00	0.00	0.08	0.07	0.00	0.00
HEK2_106	chr12	54405073	0.2	GAACAacAAGgATAaACTGCCGG	0.21	0.21	0.06	0.03	0.21	0.15
HEK2_110	chr20	17801623	0.2	GAAtgCcAAGCATAaACTGCAGG	0.00	0.00	0.04	0.02	0.00	0.00
HEK2_119	chr19	32225924	0.2	GgACACAgAGtATAGACTGaGGG	0.00	0.00	0.05	0.05	0.00	0.00
HEK2_121	chr13	58440671	0.1	GAAttCAAAGCATgGACTGCAGG	0.00	0.02	0.05	0.10	0.06	0.07
HEK2_139	chr7	113235709	0.1	tAAaACAAAGaAcAGACTGCTGG	0.11	0.09	0.02	0.03	0.11	0.11
HEK2_145	chr14	89446986	0.1	tActAtggtGCATAGACTGCTGG	0.00	0.00	0.00	0.00	0.00	0.00

RNF2

					Base editing frequency (%)				Indel frequency (%)	
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
RNF2-1	chr1	210858376	7.3	tcaccaTTAGTCATTACCTGCTG	0.03	0.02	0.14	0.09	0.00	0.01
RNF2-2	chr1	185056773	4.4	GTCATCTTAGTCATTACCTGAGG	0.02	26.26	0.05	78.82	0.01	88.29
RNF2-3	chr17	53928598	4.3	AGTCATCTTAGTCATTAC-TGAGG	0.05	0.07	0.29	0.25	0.00	0.01
RNF2-4	chr6	143212078	4.1	GTaATaTTAGTCATTACCgGTGG	0.05	0.05	0.00	0.00	0.01	0.01
RNF2-5	chr17	17072075	3.6	GTCATCcTAGTCATTtaCTGGGG	0.06	0.07	0.10	0.12	0.00	0.01
RNF2-6	chr12	131619804	2.2	aTCAcCTTAGcCATTACCaGGGG	0.06	0.09	0.09	0.10	0.02	0.02
RNF2-7	chr9	102127746	1.4	GTCA-CTTAGTCATTgCCTGTGG	0.13	0.08	0.05	0.07	0.01	0.01
RNF2-8	chr15	92315580	1.0	GTCAcaTTAGcCATTACCTGTGA	0.08	0.10	0.04	0.05	0.01	0.00
RNF2-15	chr2	51894204	0.4	aTCATCTTcaTCATTACaTGAGG	0.02	0.01	0.11	0.09	0.01	0.02
RNF2-19	chr4	125521739	0.4	ttaATCTTAGTCATTACtTtTGG	0.00	0.02	0.06	0.06	0.07	0.09
RNF2-20	chr2	102668931	0.3	aTCATCaTcGTCATTAtCTGGGG	0.08	0.12	0.06	0.05	0.04	0.04
RNF2-24	chr1	111785389	0.3	tcCATCTCAcTCATTACCTGAGG	0.03	0.03	0.02	0.05	0.29	0.35
RNF2-27	chr6	5869239	0.3	GcCAcCTcAGTCATTAgCTGGGG	0.02	0.03	0.06	0.06	0.01	0.00
RNF2-41	chr18	62061226	0.2	aTtATtTTAGTCATTACCTtTGG	0.03	0.01	0.00	0.00	0.03	0.03

(Continued)

TYRO3

	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Base editing frequency (%)			Indel frequency (%)	
					Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
TYRO3-1	chr15	76553332	4.3	GGCCACACTAGtGTTGCcGCTGG	0.30	7.77	0.08	3.84	0.06	12.94
TYRO3-2	chr15	41857272	3.4	GGCCACACTAGCGTTGCTGCTGG	0.48	71.60	0.07	39.78	0.05	49.28
TYRO3-3	chr20	45312757	2.0	GGCCACACCaAGCcTTGCTGtCGG	3.38	3.51	0.18	0.27	0.00	0.14
TYRO3-4	chr5	173133468	0.8	GGCCAC--TAGCGTTGCTcCAGG	0.26	0.21	0.15	0.16	0.00	0.00
TYRO3-5	chr22	50686658	0.6	GGCCACACTGaCcTTGCTGCTGG	0.50	0.38	0.09	0.10	0.00	0.02
TYRO3-6	chr9	140109027	0.4	ttCCACAC-AGCtTTGCTGCTGG	0.25	0.24	0.08	0.06	0.01	0.00
TYRO3-7	chr17	17207899	0.3	GGCCAC-CTAGgGTTGCTGCTGG	0.27	0.27	0.51	0.38	0.01	0.01
TYRO3-8	chr11	99666699	0.2	ccaCcCtCTAGCaTTGCTGCTGG	0.00	0.00	0.26	0.19	0.00	0.00
TYRO3-9	chr11	27495183	0.2	cctaACACCcAaCGTTGCTGCTGG	0.26	0.23	0.06	0.10	0.00	0.00
TYRO3-10	chr1	38684664	0.2	ctgagCtCTAGCaTTGCTGCTGG	0.08	0.10	0.05	0.02	0.00	0.01
TYRO3-11	chr3	22936710	0.1	GGCatataTAGCaTTGCTGCTGG	0.07	0.11	0.00	0.00	0.00	0.00
TYRO3-12	chr6	150629124	0.1	GaaCACACTAaCaTTGCTGtGGG	0.33	0.35	0.19	0.16	0.00	0.00

WEE1

	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Base editing frequency (%)			Indel frequency (%)	
					Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
WEE1-1	chr1	234241713	11.1	tTaGAGATGTTCTATTA-TCCGG	0.26	0.26	0.00	0.00	0.00	0.06
WEE1-2	chr11	9610086	2.0	GTCGAGATGTTCTATTACTCTGG	0.05	62.59	0.11	9.95	0.00	56.24
WEE1-3	chr4	18855558	1.9	GagGAGATGcTCTATTACTCCGG	0.38	0.42	0.00	0.00	0.00	4.13
WEE1-4	chrX	42756785	0.7	aTCaAGATGaaCTATTACTCTGG	0.13	0.11	0.06	0.03	0.00	0.01
WEE1-5	chr1	215805743	0.4	GTgGAGATGTTaTgTTACTCTGG	0.08	0.11	0.00	0.00	0.00	0.01
WEE1-6	chr5	57863350	0.4	GaCataATGTTCTATTACTCAAG	0.10	0.10	0.06	0.05	0.00	0.01
WEE1-7	chr4	119010382	0.3	taatAGATGTTCTATTACTaAGG	0.04	0.06	0.00	0.00	0.00	0.02
WEE1-8	chr15	80845387	0.3	aTtaAGATGTTCTATTAaTaGGG	0.25	0.24	0.00	0.00	0.01	0.00
WEE1-9	chr13	87867165	0.2	cagctaATGTaaTATTACTCAGG	0.05	0.04	0.12	0.17	0.01	0.01
WEE1-10	chr4	107037600	0.2	aTCaAGATaTTCTATTACTgGGG	0.11	0.07	0.17	0.15	0.04	0.05
WEE1-11	chrX	136297269	0.2	tctrAtATtTTCTATTACTCAGG	0.24	0.29	0.00	0.00	0.46	0.43
WEE1-12	chr1	8500986	0.2	aTaGAGCaGTTaTATTACTCTGG	No PCR					
WEE1-13	chr11	71806684	0.2	GTaGAGATGTTCT-TaACTCAGG	0.67	0.69	0.00	0.00	0.40	0.29
WEE1-14	chr12	83599167	0.2	cTCttttatTTCTATTACTCTGG	0.00	0.00	0.13	0.13	0.00	0.01
WEE1-15	chr8	71394697	0.1	aatataATaTTCTATTACTCAGG	0.14	0.11	0.00	0.00	0.01	0.01
WEE1-16	chr8	86942319	0.1	caactataaTTCTATTACTCTGG	0.00	0.03	1.37	1.29	0.00	0.00
WEE1-17	chr1	61165563	0.1	GgCtActgtcTCTATTACTCAGG	0.14	0.08	0.13	0.20	0.02	0.00
WEE1-18	chr20	45730192	0.1	aTttAGATaTTCTATTACTCTGA	0.09	0.11	0.00	0.00	0.00	0.01
WEE1-19	chr8	134795474	0.1	cTttAagataTCTATTACTCTGG	0.11	0.12	0.00	0.00	0.01	0.00
WEE1-20	chr2	128238175	0.1	caaGAGAcGcTCTATTcACTCAGG	0.26	0.25	0.00	0.00	0.00	0.00
WEE1-21	chr7	31677975	0.1	aaacAGcTGTTCTATTACTCAGG	0.09	0.12	0.27	0.19	0.02	0.00

(Continued)

EphB4

	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Base editing frequency (%)				Indel frequency (%)	
					Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
EphB4-01	chr7	100411300	10.3	GCAGAATATTCGGACAAACACGG	0.12	57.62	0.14	0.17	0.01	74.73
EphB4-02	chr5	81021421	8.0	GCAGAATATTaCaGACAAACtcAGG	0.15	0.23	0.00	0.00	0.00	1.02
EphB4-03	chr15	47712791	4.1	GCAGAATATTcaGGACAAACAATG	0.09	0.11	0.26	0.30	0.01	0.14
EphB4-04	chr11	123329727	2.4	aCAGAATATTCaGACAAtCACAG	0.09	0.10	0.05	0.06	0.01	0.01
EphB4-05	chr1	93047514	2.2	GCAGtA-AaTaaGACAAACAAGG	0.14	0.13	0.08	0.09	0.01	0.00
EphB4-06	chr15	51198151	1.9	aCAaAATATTCGGACAggCACGG	No PCR					
EphB4-07	chr18	29712895	1.8	GCAGAATtTaaGGACAAgCAAGG	0.33	0.30	0.09	0.10	0.00	0.01
EphB4-08	chr3	55670295	1.8	GCAGAATA--aGGACAAACATGG	0.12	0.15	0.07	0.14	0.00	0.00
EphB4-09	chr13	25919264	1.4	cagtAAATATTCaGACtAAACATGG	0.50	0.42	0.00	0.00	0.01	0.01
EphB4-10	chr3	122057177	0.9	ttAatcATATaaGGACAAACATGG	0.07	0.08	0.00	0.00	0.01	0.00
EphB4-11	chr1	26891812	0.8	aCAtAATATTCaaACAAACAGGG	0.07	0.08	0.16	0.16	0.01	0.02
EphB4-12	chr10	131074805	0.6	aatGAATAaagGGACAAACAGGG	0.52	0.54	0.00	0.00	0.01	0.01
EphB4-13	chr4	123425007	0.5	tCAG-ATATaaGGACAAACATGG	0.15	0.14	0.11	0.09	0.01	0.01
EphB4-14	chr5	60692862	0.4	GCtGAATATTgGGACAAtCATGG	0.15	0.14	0.10	0.13	0.01	0.00
EphB4-15	chr9	930844	0.4	GCAGtcTATAaaaaGACAAACAGGG	0.12	0.10	0.00	0.00	0.01	0.01
EphB4-16	chr9	29344386	0.4	aCAGAATATaaGaACAAAtcATGG	0.06	0.08	0.13	0.11	0.01	0.00
EphB4-17	chr4	160752620	0.4	atAGAATAaTCaGACAAAaAAGG	0.20	0.20	0.00	0.00	0.32	0.33
EphB4-18	chr16	23706770	0.4	GCAG-ATAaagGGACAAACATGG	0.60	0.69	0.35	0.30	0.00	0.01
EphB4-19	chr11	117354211	0.3	GacagtgAaTaGGACAAACATGG	0.11	0.09	0.10	0.10	0.00	0.00
EphB4-20	chr14	90497703	0.3	tgAGAATATgCGGACAAACACAG	0.23	0.20	0.00	0.00	0.98	1.10
EphB4-21	chr2	68459706	0.3	atAGAATATTaGGACAAAtAGAG	No PCR					
EphB4-22	chr13	51951355	0.2	caAtAATAaaaGGACAAACATGG	0.04	0.04	0.00	0.00	0.00	0.00
EphB4-23	chr17	7777056	0.2	aCAGgATAaaaGGACAAAtATGG	0.05	0.06	0.05	0.05	0.12	0.11
EphB4-24	chr3	188431448	0.2	aCAGAtATAcgtaGACAAACAGAG	0.27	0.31	99.98	99.98	0.00	0.01
EphB4-25	chr12	11705824	0.2	atAGAATAagaaGACAAACATGG	0.06	0.06	0.00	0.00	0.00	0.00
EphB4-26	chr10	95697506	0.2	ttAGAATATTCcttACAAACAAGG	0.13	0.10	0.00	0.00	0.00	0.00
EphB4-27	chrX	123665071	0.2	cCAtgtTATTCaGACAAACATGG	No PCR					
EphB4-28	chr16	75869847	0.2	GCAGAA-AaTaGGACAAtCACGG	0.29	0.27	0.12	0.13	0.00	0.00
EphB4-29	chr6	128784409	0.2	agAGAATATa-GGACAAACtcTGG	0.55	0.48	0.00	0.00	0.00	0.01
EphB4-30	chr5	144297821	0.2	GCAGAAAataTttGACAAACAAGG	0.10	0.08	0.03	0.04	0.00	0.00
EphB4-31	chr14	29760524	0.1	aaAGAATATa-GaACAAACAGGG	0.15	0.18	0.00	0.00	0.24	0.21
EphB4-32	chr1	36914336	0.1	aCAGAA-ATaCGaACAAACATGG	0.54	0.45	0.03	0.01	0.00	0.00
EphB4-33	chr21	26227853	0.1	aacaAATgcaaGGACAAACATGG	0.23	0.15	0.15	0.11	0.03	0.01
EphB4-34	chr13	63528393	0.1	aCActcAT-TTaaGACAAACAGGG	0.15	0.13	0.21	0.27	0.01	0.00
EphB4-35	chr10	72516747	0.1	aCAGActacgaGACAAACACGG	6.14	6.02	0.20	0.31	0.00	0.00
EphB4-36	chr7	20108292	0.1	GCAGA-TAccCaGACAAACAGGG	No PCR					
EphB4-37	chr1	89896856	0.1	taAGAATAT-aGGACAAACAGGA	0.22	0.19	0.00	0.00	0.00	0.00
EphB4-38	chr4	145533104	0.1	cCAGAATATTa-GACAAACATGG	0.12	0.12	0.43	0.40	0.11	0.13
EphB4-39	chr20	9784096	0.1	agAGAATATagGGACAAAtAAGG	0.11	0.11	0.00	0.00	0.01	0.00

(Continued)

HPRT-Exon6

	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Base editing frequency (%)			Indel frequency (%)	
					Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
HPRT_E6-01	chrX	133627607	29.0	GTATAATCCAAAGATGGTCAAGG	0.11	62.07	0.05	29.25	0.00	71.69
HPRT_E6-02	chr12	7468915	21.3	GTATAcatCcAAGATGGcCAGGG	0.22	0.21	0.01	0.01	0.00	0.00
HPRT_E6-03	chr3	10057621	17.0	GTATtA-CCAAAGATGGTCtGGG	0.08	0.18	0.03	0.01	0.00	0.00
HPRT_E6-04	chr22	39586062	12.7	GTATAATC-AAAGATGGcCcTGG	0.38	0.40	0.00	0.00	0.00	0.00
HPRT_E6-05	chr7	13928792	10.5	GTATtAT-CAAAaATGGTCAAGG	0.15	0.14	0.00	0.00	0.00	0.00
HPRT_E6-06	chr1	225831201	10.2	aTATAA-CCAAAGATGTtTCACAG	0.30	0.31	0.59	0.44	0.00	0.00
HPRT_E6-07	chr17	53575925	9.3	aTATAATCtcAAGATGGTCAAGG	0.02	1.59	0.28	0.28	0.81	17.56
HPRT_E6-08	chr12	13406973	9.0	aTATAATCCtAAGATGtTCATGG	0.12	0.12	0.04	0.03	0.00	0.17
HPRT_E6-09	chr14	27255077	6.1	GTATtATCCAAAGATGacCAGAG	0.07	0.05	0.04	0.06	0.00	0.00
HPRT_E6-10	chr7	85914581	5.6	aTAcAATCCAAAGtTGGTCtGGG	0.24	0.14	0.13	0.12	0.00	0.01
HPRT_E6-11	chr3	9906633	4.7	aTATAttaCCAAAGATGGTCtGGG	0.15	0.17	0.01	0.02	0.00	0.01
HPRT_E6-12	chr8	125496169	4.5	GTAgAATCCAtAGATGGaCAGGG	0.26	0.31	0.07	0.06	0.00	0.27
HPRT_E6-13	chr17	10064482	4.3	GTATcA-CCAAAGATGGTCATGG	0.35	1.89	0.05	0.29	0.00	0.00
HPRT_E6-14	chr2	53374352	4.3	GTAcAATCaAAAGATGaaCAAGG	0.10	0.11	0.15	0.17	0.00	0.00
HPRT_E6-15	chr4	102026801	4.1	aTATtATCCAAAGATGGagtTGG	0.07	0.06	0.11	0.07	0.00	0.00
HPRT_E6-16	chr6	19070067	3.8	aTgTAtTCCAAAGATGGTCAGGG	0.10	0.34	0.41	1.74	0.01	0.00
HPRT_E6-17	chr6	7761387	3.7	aagTAATCCAAAGATaGTCtTGG	0.08	0.09	0.10	0.08	0.00	0.00
HPRT_E6-18	chr12	59613203	3.6	GTATAAcCCAAAGAaaGTaATGG	0.26	0.31	0.36	0.39	0.00	0.00
HPRT_E6-19	chr16	20336811	3.6	aTATAtATtCcAAGATGGTtcTGG	0.06	0.05	0.00	0.00	0.00	0.00
HPRT_E6-20	chr4	136809212	3.5	tTATAATCCAAAGAaGGcCAAGG	0.00	0.00	0.05	0.08	0.00	0.15
HPRT_E6-21	chr12	117915656	3.5	GTATctAcCCAAAGATGtTCATGG	0.07	0.08	0.30	0.32	0.03	0.03
HPRT_E6-22	chr14	45005861	3.4	GTATAAT--AAAGATGGTCAAGG	0.16	0.21	0.00	0.00	0.00	0.01
HPRT_E6-23	chr14	104033102	3.3	aTAgAtTCCAAAGATGGaCATGG	0.33	0.22	0.07	0.08	0.00	0.01
HPRT_E6-24	chr11	93731924	3.1	cTATAATCtAAAAaTGGTCAAGG	0.09	0.06	0.06	0.05	0.00	0.01
HPRT_E6-25	chr13	75257165	2.9	GTATAcTagAAAGATGGTCATGA	0.13	0.10	0.00	0.00	0.01	0.00
HPRT_E6-26	chr15	54911620	2.8	GgATAATCaAAAGATGGaCtTGG	0.11	0.07	0.04	0.07	0.00	0.00
HPRT_E6-27	chrX	6849511	2.6	aTATAATgCAAAGcTGGTCACGG	0.22	0.27	0.00	0.00	0.00	0.00
HPRT_E6-28	chr11	87526371	2.5	aTAcAATCCAAAGATatTCtAGG	0.33	0.29	0.27	0.20	0.00	0.01
HPRT_E6-29	chr3	19652583	2.4	GTATAATCCAAAGgTGGcCtAGG	0.23	0.24	0.04	0.03	0.07	0.09
HPRT_E6-30	chr17	45009128	2.4	aTATAATCCAtAGaAGGTCAAGA	0.50	0.49	0.03	0.04	0.00	0.00
HPRT_E6-31	chr8	9090663	2.3	GatatAatacAAGATGGaCAAGG	0.21	0.24	0.00	0.00	0.00	0.00
HPRT_E6-32	chr12	6810501	2.2	GgATAATCCAAAGATGGTgcAAG	0.24	0.29	0.05	0.07	0.01	0.00
HPRT_E6-33	chr6	118233456	1.5	aTATAAaCCAAAGATcGTtAAGG	0.22	0.23	0.08	0.07	0.01	0.01
HPRT_E6-34	chr14	85667729	1.4	tctcccTCCAAAGAaGGTCACGG	0.00	0.00	0.63	0.54	0.00	0.00
HPRT_E6-35	chrX	110551547	1.4	tTATAgTCCAAAGgaaGTCAGGG	0.20	0.22	0.02	0.05	0.00	0.00
HPRT_E6-36	chr8	18630339	1.2	tatTtATCCAAAGATGGTCtAGA	0.05	0.07	0.05	0.03	0.15	0.10
HPRT_E6-37	chr12	45338334	1.1	aTATAtaCCAAAGgTaGTCATGG	0.02	0.01	0.07	0.07	0.00	0.00
HPRT_E6-38	chr7	135056752	1.1	aggagATCCAAAGATGGTCAAGG	0.16	0.25	0.06	0.89	0.01	0.00
HPRT_E6-39	chr9	82090375	1.1	acATAgTCCAAAtATGGaCAGGG	0.04	0.04	0.11	0.12	0.00	0.00

(Continued)

HPRT_E6-40	chr12	103187212	1.0	GTATAtAattcAAGATGGTCAGGA	0.12	0.09	0.00	0.00	0.00	0.00	0.00	0.01
HPRT_E6-41	chr12	58996606	1.0	GaATAAgCCAAAGATGGTCAGTG	0.25	0.30	0.00	0.00	0.00	0.00	0.00	0.00
HPRT_E6-52	chr12	29153372	0.7	GTATttTCCAAAcGATGGTCATGG	0.00	0.00	0.01	0.06	0.01	0.01	0.00	0.00
HPRT_E6-81	chr11	44024764	0.4	GTATAAgCCAAAGcaGGTCACGG	0.21	0.23	0.08	0.08	0.00	0.00	0.00	0.00

HPRT-Exon8

					Base editing frequency (%)				Indel frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	Untreated	(+) BE3	Untreated	(+) Cas9
HPRT_E8-1	chrX	133632686	2.4	GAAGTATTCATTATAGTCAAGGG	0.02	70.04	0.05	27.31	0.07	64.71
HPRT_E8-2	chr11	93382674	1.3	GAAGcATTCATTATAGTCAAAGG	0.20	5.77	0.09	0.58	0.07	50.40
HPRT_E8-3	chr21	34111536	0.6	GAAGaATTCATTATAGaCAATGG	0.16	0.16	0.05	0.02	0.01	0.04
HPRT_E8-4	chr11	100880874	0.6	aAgcTATTCATTATAGcaAATGG	0.04	0.06	63.15	63.00	0.01	0.01
HPRT_E8-5	chr22	29285654	0.6	GAAGTaATTCATTATAGaCAATGG	0.10	0.08	0.10	0.11	0.00	0.02
HPRT_E8-6	chr4	15866630	0.5	aAAGTATTaATTATAGTCAAGGA	0.03	0.07	N/A	N/A	0.01	0.19
HPRT_E8-7	chr2	53543615	0.5	GAAGTAT-CcTTATAGTCAgAAG	0.16	0.15	0.07	0.11	0.00	0.00
HPRT_E8-8	chr3	149876617	0.4	GAAGTAT-CATTATAGTCtttGGG	0.07	0.07	0.08	0.05	0.03	0.04
HPRT_E8-9	chr8	71429233	0.4	GcAGTATTCATaATAGTCAAAAG	0.02	0.03	0.06	0.09	0.02	0.04
HPRT_E8-10	chrX	146032194	0.4	aAAGTATTTcaTATAGcCAAAGG	0.01	0.01	N/A	N/A	0.01	0.03
HPRT_E8-11	chr5	132311070	0.4	GAAtTATTCATcATAGcCAAAGG	0.08	0.07	0.09	0.09	0.04	0.03
HPRT_E8-12	chr3	36349740	0.3	tAAGTAaTCATTATAGTCAgATG	0.10	0.09	0.05	0.05	0.00	0.01
HPRT_E8-13	chr3	181655989	0.3	atAGTATTCATTATAGTaAcAGG	0.07	0.03	0.06	0.05	0.00	0.03
HPRT_E8-14	chr11	116029576	0.3	aAAaTATcCATTATAGTCttACG	0.08	0.14	0.67	0.48	0.02	0.00
HPRT_E8-15	chr10	131905650	0.3	GAAcTATTCATTATtaTCAACGG	0.10	0.07	0.09	0.07	0.00	0.01
HPRT_E8-16	chr9	89312308	0.2	GttGTATTCATTATAGTCAAGAG	0.02	0.02	0.08	0.09	0.02	0.05
HPRT_E8-17	chr4	178543184	0.2	atAGTATTCATTATAGTaAcAGG	0.10	0.11	0.05	0.05	0.00	0.02
HPRT_E8-18	chr12	72733535	0.2	agAGTATcCAccATAGTCAAGGG	0.05	0.08	0.14	0.10	0.01	0.20
HPRT_E8-19	chr1	45916291	0.1	GAAGTAcatcTTATAaTCAATGG	0.05	0.06	0.06	0.03	0.03	0.01
HPRT_E8-20	chr4	153599265	0.1	aAtGTATTCATTATAGgCAAGAG	0.15	0.17	0.04	0.04	0.01	0.00

(With Daesik Kim in Institute for Basic Science)

5. Increasing ABE7.10 genome-wide specificities using modified sgRNAs, Sniper-Cas9 and Ribonucleoprotein delivery

To minimize or avoid ABE off-target activity, three different methods were used, proven useful for reducing Cas9 and BE3 off-target effects: sgRNA modification (Cho et al., 2014; Fu et al., 2014; Kim et al., 2015; Kim et al., 2016; Tsai et al., 2015), use of Sniper-Cas9 (Lee et al., 2018b), an engineered Cas9 variant obtained via directed evolution in *E. coli*, and delivery of RNPs in lieu of plasmids (Kim et al., 2017b; Kim et al., 2014; Rees et al., 2017),

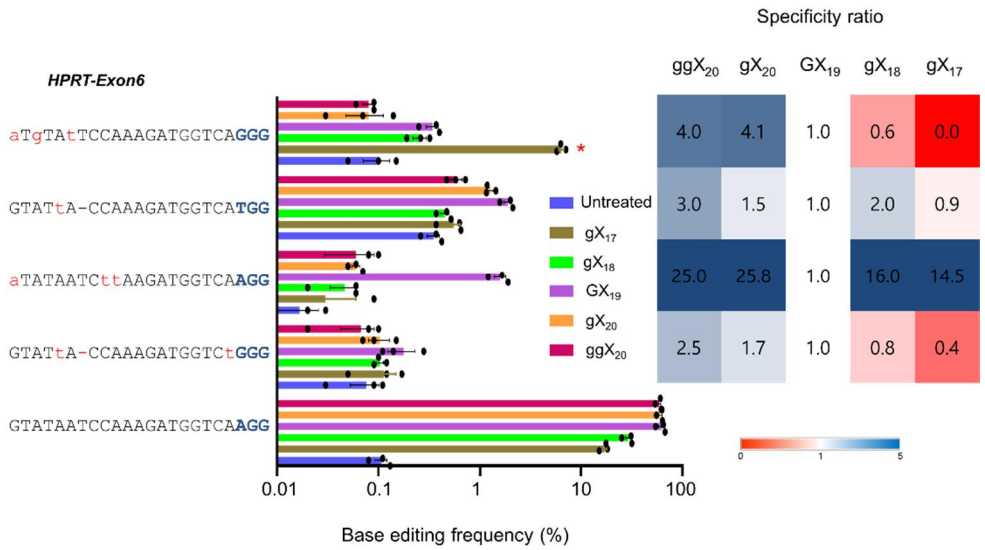
First, I replaced standard GX_{19} sgRNAs with truncated sgRNAs (termed gX_{18} or gX_{17}) ("g" and "G" represent, respectively, a mismatched and matched guanine) or extended sgRNAs containing one or two extra guanines at the 5′ terminus (termed gX_{20} or ggX_{20}) and measured editing frequencies in HEK293T cells (Figure 15a and Figure 16a). Extended sgRNAs reduced ABE off-target activity at almost every site without sacrificing its on-target activity (Figure 16b, Figure 16b, and Table 7). Thus, at the four validated off-target sites identified using a GX_{19} sgRNA targeted to *HPRT* exon 6, gX_{20} or ggX_{20} sgRNAs showed 2 to 26-fold improvement in specificity ratios, compared to the conventional GX_{19} sgRNA (Figure 15b). By contrast, truncated sgRNAs reduced both ABE on-target and off-target activity at most sites and exacerbated off-target effects at sites with mismatches at or near the 5' terminus (shown by asterisks and heatmaps in Figure 15b and Figure 16b).

I also found that use of Sniper ABE7.10, in which Sniper-Cas9 in lieu of the wild-type SpCas9 was fused to the adenine deaminase moiety (Figure 17 and Table 8), or combining Sniper ABE7.10 with modified sgRNAs further reduced ABE off-target activity at many validated off-target sites (Figure 18). Delivery of ABE7.10 RNPs rather than plasmids improved the specificity of base editing by a factor of up to 7 (Figure 19).

a

5′-GTATAATCCAAAGATGGTCA**NGG**-3′	*HPRT exon 6* target sequence
5′-gTAATCCAAAGATGGTCA-3′	gX_{17} sgRNA
5′-gATAATCCAAAGATGGTCA-3′	gX_{18} sgRNA
5′-GTATAATCCAAAGATGGTCA-3′	GX_{19} sgRNA
5′-gGTATAATCCAAAGATGGTCA-3′	gX_{20} sgRNA
5′-ggGTATAATCCAAAGATGGTCA-3′	ggX_{20} sgRNA

b

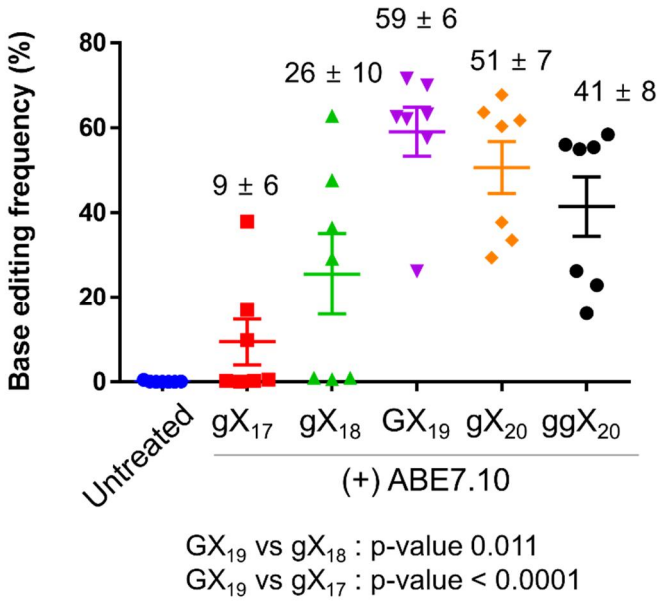


(With Daesik Kim in Institute for Basic Science)

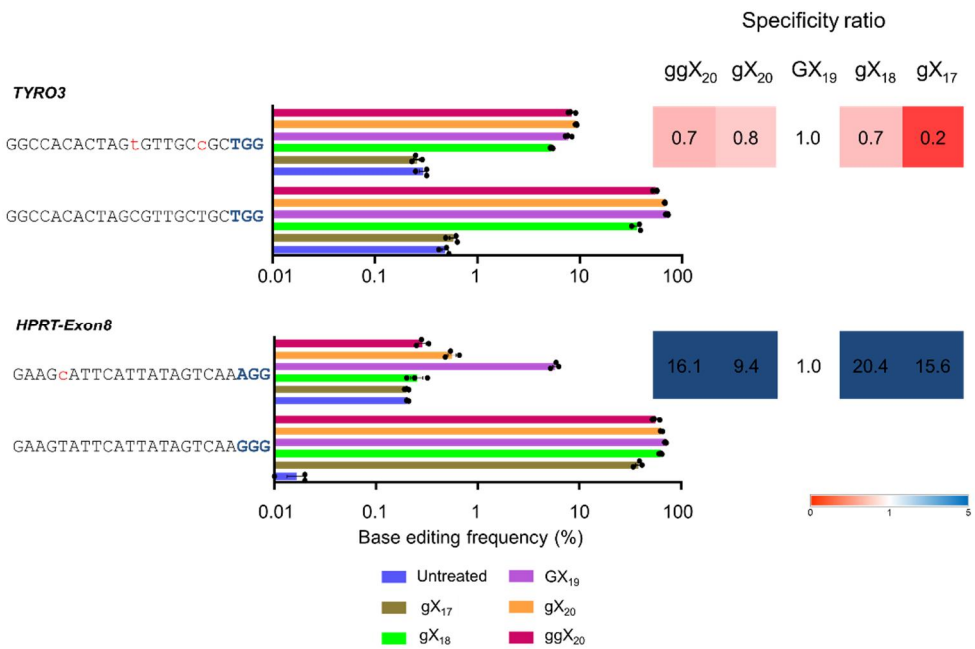
Figure 15. Reducing ABE7.10 off-target effects with modified sgRNAs. (a) Target DNA-complementary RNA sequences of conventional (GX_{19}), truncated (gX_{18} or gX_{17}), and extended (gX_{20} or ggX_{20}) sgRNAs. (b) ABE7.10-mediated base-editing frequencies determined by targeted deep sequencing at the *HPRT-Exon 6* on- or off-target sites in HEK293T cells. The heatmap shows relative specificities of modified sgRNAs, compared to that of the GX_{19} sgRNA.

Specificity ratios were calculated by dividing the specificity of sgRNA of interest (on-target frequency/off-target frequency) by the specificity of the GX_{19} sgRNA (on-target frequency/off-target frequency). Means ± s.e.m. were from three independent experiments.

a



b



(With Daesik Kim in Institute for Basic Science)

Figure 16. On- and off-target activity measured using modified sgRNAs. (a) Mean of base editing frequency when using modified sgRNAs targeted to *HEK2, RNF2, TK_EphB4, TYRO3, WEE1, HPRT*-exon6, and *HPRT*-exon8. All data are the results of targeted deep sequencing. (b) Off-target activities measured using modified sgRNAs at the *TYRO3* and *HPRT*-exon8 sites. The red and blue characters indicate the mismatched nucleotides and PAM sequences, respectively. The heatmap shows relative specificities of modified sgRNAs, compared to that of the GX_{19} sgRNA. The specificity ratio was calculated using the following formula: specificity of modified sgRNAs (on-target frequency/off-target frequency)/ specificity of conventional sgRNAs (on-target frequency/off-target frequency). Means ± s.e.m. were from three independent experiments.

Table 7. Analysis of ABE7.10 off-target effects via modified sgRNAs.

HEK2					Base editing frequency (%)							
						Untreated	(+) ABE7.10					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites			gX17	gX18	GX19	gX20	ggX20
HEK2_001	chr19	35505485	20.1	GAACAC−AAGCAcAGACTGAaAGG		0.12	0.22	0.15	0.11	0.12	0.13
HEK2_002	chr15	93557679	15.1	GAACACA−tGCATAGACTGCTAG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_003	chr4	53536209	14.7	GAAtACtAAGCATAGACTcCAGG		0.06	0.07	0.04	0.09	0.08	0.07
HEK2_004	chr4	90522183	14.3	GAACACAAtGCATAGAtTGCCGG		0.10	0.10	0.11	0.31	0.21	0.11
HEK2_005	chr1	167742859	13.4	aAACACAgAGCAcAGACTGCTGA		0.15	0.14	0.16	0.13	0.13	0.12
HEK2_006	chr10	43159340	13.4	aAACACAAAGaCATAGACcaCTGG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_007	chr5	87240613	8.9	GAACACAAAGCATAGACTGCGGG		0.06	9.80	47.71	63.19	61.77	55.40
HEK2_008	chr9	290167	6.6	aAACAtAAAGaATAGACTGCAAG		0.06	0.07	0.05	0.12	0.12	0.08
HEK2_009	chr3	28560566	5.8	GAACACAAAGtcTAGAaTGCTAG		0.10	0.08	0.08	0.09	0.11	0.12
HEK2_010	chr1	175310138	5.1	GgACACAAAGCcTAGACTcCAGG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_011	chr13	87616351	4.9	GAAgAtAAAGCATAGACTctAGG		0.09	0.10	0.07	0.08	0.09	0.09
HEK2_012	chr10	87125365	4.6	aAACAtAAAGCATAGACTGCaAAG		0.05	0.05	0.03	0.06	0.04	0.04
HEK2_013	chr18	32279782	3.9	aAAtACAAAGCATAGACTaatATG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_014	chr2	192248363	3.5	GAACACAtA−CATAGACaGCTGG		0.08	0.09	0.13	0.08	0.05	0.07
HEK2_015	chr5	160202385	3.0	GAAaaCAAAGCAaAGAaaGCAGG		0.11	0.09	0.14	0.13	0.13	0.14
HEK2_016	chr15	63236612	2.9	aAcacaAtAGCATAGACTGgaCTG					No PCR		
HEK2_017	chr20	23101377	2.7	atACACAgAGCAaAGACTGCAGG		0.13	0.07	0.15	0.07	0.12	0.10
HEK2_018	chrX	48317375	2.5	aAACAtAAAGCcTAGACTGaCGG		0.18	0.25	0.21	0.20	0.23	0.23
HEK2_019	chr19	49972515	2.2	aAAaACAAAGCgTAGACTGtGGG		0.12	0.13	0.13	0.12	0.11	0.12
HEK2_020	chr12	81785997	2.1	GgACAaAAAGCATAGACTcaAGG		0.12	0.20	0.11	0.20	0.14	0.14
HEK2_021	chr1	40461424	2.1	atACACAcAAGCAcAGACTGCAGG		0.17	0.18	0.18	0.19	0.21	0.20
HEK2_022	chr15	92968604	2.1	GAACACAgcaCATAaACTGCAGG		0.13	0.10	0.09	0.11	0.14	0.13
HEK2_023	chr5	142142325	2.0	aAACACtAA−CATAGACTGCAGG		0.09	0.13	0.05	0.09	0.12	0.08
HEK2_024	chr5	131174461	1.7	aAAtACAAtGCATAGACTGCTAG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_025	chr1	36097072	1.4	GtAaACAAAGCATAGACTGaGGG		0.13	0.14	0.13	0.14	0.11	0.13
HEK2_026	chrX	132724931	1.3	GAACA−tAAtCAcAGACTGCTGG		0.13	0.12	0.15	0.16	0.11	0.16
HEK2_027	chr6	139353017	1.3	ccAaACAAAAaCATAGACTGCTGG		0.15	0.19	0.21	0.16	0.18	0.18
HEK2_028	chrX	36949816	1.3	GAAaACAAAaCATAGAgTGCTGG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_029	chr17	26701101	1.2	aAtCAaAtAGCATAGACTGCATG		0.23	0.20	0.16	0.15	0.18	0.14
HEK2_030	chr13	91254877	1.2	GAACACAAAcatagtACTGaAGG		0.09	0.11	0.10	0.11	0.08	0.09
HEK2_031	chrX	138877515	1.2	aAACAgAAAGCATgGACTGCGGA		0.09	0.12	0.10	0.12	0.08	0.09
HEK2_032	chr13	113428466	1.1	cAAtACAAAGgATAGACTGCAGG		0.05	0.07	0.12	0.09	0.10	0.06
HEK2_033	chr15	41796089	1.1	agACACAcAGCAcAGACTGCAGG		0.11	0.11	0.14	0.22	0.14	0.16
HEK2_034	chr11	128508576	1.0	GAAttCAAAGCATAGAtTGCAGG		0.05	0.05	0.06	0.05	0.05	0.05
HEK2_042	chr3	33744669	0.7	GAAAaACAAAcCAcAGACTGggGGG		0.03	0.07	0.04	0.04	0.05	0.04
HEK2_044	chr16	64198277	0.7	GAAtACAtAaCATAGACTGgGGG		0.00	0.00	0.00	0.00	0.00	0.00
HEK2_050	chr20	97640	0.5	GAAttCAAAGCATAGAtTGCAGG		0.22	0.23	0.13	0.19	0.16	0.19
HEK2_052	chr3	10835885	0.5	aAACACAtAGCAcAGcCTGCAGG		0.13	0.10	0.09	0.11	0.07	0.11
HEK2_065	chr18	56307002	0.4	aAgaACAAAaCATAGACTGCAGG		0.11	0.12	0.07	0.15	0.14	0.12

(Continued)

HEK2_079	chr5	126385455	0.3	ccACACcAAGCATAGACTtCTGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_082	chr12	131431445	0.3	GgAgAgAgAGCATAGACTGCTGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_092	chr1	237916998	0.2	tAACACAAAGacTAaACTGCAGG	0.11	0.08	0.11	0.08	0.06	0.06	0.06	0.10
HEK2_093	chr1	215323272	0.2	tAcCAGgAAGCATAGACTGtAGG	0.02	0.00	0.02	0.03	0.03	0.03	0.03	0.04
HEK2_099	chr13	51093353	0.2	GAACACAcA-CAcAGACTGaGGG	0.06	0.15	0.13	0.08	0.10	0.10	0.10	0.13
HEK2_101	chr1	20011125	0.2	aAAgctAcAGCATAGACTGCAAG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_105	chr19	5068807	0.2	tAtCAgAAAGCAcAGACTGCGGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_106	chr12	54405073	0.2	GAACAacAAGgATAaACTGCCGG	0.21	0.23	0.26	0.21	0.27	0.27	0.27	0.19
HEK2_110	chr20	17801623	0.2	GAAtgCcAAGCATAaACTGCAGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_119	chr19	32225924	0.2	GgACACAgAGtATAGACTGaGGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HEK2_121	chr13	58440671	0.1	GAAtCAAAGCATgGACTGCAGG	0.00	0.02	0.04	0.02	0.07	0.07	0.07	0.02
HEK2_139	chr7	113235709	0.1	tAAaACAAAGaAcAGACTGCTGG	0.11	0.12	0.09	0.09	0.17	0.17	0.17	0.14
HEK2_145	chr14	89446986	0.1	tActAtggtGCATAGACTGCTGG	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

RNF2

					Base editing frequency (%)						
						(+) ABE7.10					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	gX_{17}	gX_{18}	GX_{19}	gX_{20}	ggX_{20}
RNF2-1	chr1	210858376	7.3	tcaccaTTAGTCATTACCTGCTG	0.03	0.01	0.02	0.02	0.01	0.03
RNF2-2	chr1	185056773	4.4	GTCATCTTAGTCATTACCTGAGG	0.02	0.07	0.63	26.26	29.43	16.19
RNF2-3	chr17	53928598	4.3	AGTCATCTTAGTCATTAC-TGAGG	0.05	0.07	0.07	0.07	0.05	0.06
RNF2-4	chr6	143212078	4.1	GTaATaTTAGTCATTACCgGTGG	0.05	0.05	0.05	0.05	0.04	0.05
RNF2-5	chr17	17072075	3.6	GTCATCcTAGTCATTtaCTGGGG	0.06	0.06	0.08	0.07	0.07	0.08
RNF2-6	chr12	131619804	2.2	aTCAcCTTAGcCATTACCaGGGG	0.06	0.08	0.04	0.09	0.06	0.05
RNF2-7	chr9	102127746	1.4	GTCA-CTTAGTCATTgCCTGTGG	0.13	0.15	0.09	0.08	0.17	0.16
RNF2-8	chr15	92315580	1.0	GTCAcaTTAGcCATTACCTGTGA	0.08	0.11	0.10	0.10	0.08	0.07
RNF2-15	chr2	51894204	0.4	aTCATCTTcaTCATTACaTGAGG	0.02	0.02	0.03	0.01	0.05	0.02
RNF2-19	chr4	125521739	0.4	tTaATCTTAGTCATTACtTtTGG	0.00	0.02	0.02	0.02	0.00	0.01
RNF2-20	chr2	102668931	0.3	aTCATCaTcGTCATTAtCTGGGG	0.08	0.10	0.10	0.12	0.10	0.10
RNF2-24	chr1	111785389	0.3	tcCATCTcAcTCATTACCTGAGG	0.03	0.01	0.02	0.03	0.03	0.03
RNF2-27	chr6	5869239	0.3	GcCAcCTcAGTCATTAgCTGGGG	0.02	0.03	0.05	0.03	0.03	0.01
RNF2-41	chr18	62061226	0.2	aTcATtTTAGTCATTACCTtTGG	0.03	0.01	0.01	0.01	0.01	0.01

(Continued)

TYRO3

					Base editing frequency (%)					
					Untreated	(+) ABE7.10				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites		gX_{17}	gX_{18}	GX_{19}	gX_{20}	ggX_{20}
TYRO3-1	chr15	76553332	4.3	GGCCACACTAGtGTTGCcGCTGG	0.30	0.26	5.32	7.77	9.26	8.32
TYRO3-2	chr15	41857272	3.4	GGCCACACTAGCGTTGCTGCTGG	0.48	0.58	36.53	71.60	67.78	54.99
TYRO3-3	chr20	45312757	2.0	GGCCACACcAGCcTTGCTGtCGG	3.38	3.44	3.40	3.51	3.34	3.53
TYRO3-4	chr5	173133468	0.8	GGCCAC--TAGCGTTGCTcCAGG	0.26	0.20	0.21	0.21	0.24	0.24
TYRO3-5	chr22	50686658	0.6	GGCCACACTgaCcTTGCTGCTGG	0.50	0.41	0.50	0.38	0.51	0.42
TYRO3-6	chr9	140109027	0.4	ttCCACAC-AGCtTTGCTGCTGG	0.25	0.25	0.24	0.24	0.26	0.25
TYRO3-7	chr17	17207899	0.3	GGCCAC-CTAGgGTTGCTGCTGG	0.27	0.31	0.25	0.27	0.31	0.27
TYRO3-8	chr11	99666699	0.2	ccaCcCtCTAGCaTTGCTGCTGG	0.00	0.00	0.00	0.00	0.00	0.00
TYRO3-9	chr11	27495183	0.2	cctaACACcAaCGTTGCTGCTGG	0.26	0.21	0.24	0.23	0.21	0.21
TYRO3-10	chr1	38684664	0.2	ctgagCtCTAGCaTTGCTGCTGG	0.08	0.06	0.07	0.10	0.05	0.08
TYRO3-11	chr3	22936710	0.1	GGCatataTAGCaTTGCTGCTGG	0.07	0.10	0.10	0.11	0.09	0.08
TYRO3-12	chr6	150629124	0.1	GaaCACACTAaCaTTGCTGtGGG	0.33	0.30	0.37	0.35	0.32	0.30

WEE1

					Base editing frequency (%)					
					Untreated	(+) ABE7.10				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites		gX_{17}	gX_{18}	GX_{19}	gX_{20}	ggX_{20}
WEE1-1	chr1	234241713	11.1	tTaGAGATGTTCTATTAtTCCGG	0.26	0.26	0.24	0.26	0.26	0.23
WEE1-2	chr11	9610086	2.0	GTCGAGATGTTCTATTACTCTGG	0.05	0.25	0.95	62.59	37.80	26.30
WEE1-3	chr4	18855558	1.9	GagGAGATGcTCTATTACTCCGG	0.38	1.80	0.40	0.42	0.42	0.39
WEE1-4	chrX	42756785	0.7	aTCaAGATGaaCTATTACTCTGG	0.13	0.08	0.09	0.11	0.11	0.11
WEE1-5	chr1	215805743	0.4	GTgGAGATGTTaTgTTACTCTGG	0.08	0.10	0.12	0.11	0.10	0.07
WEE1-6	chr5	57863350	0.4	GaCataATGTTCTATTACTCAAG	0.10	0.12	0.12	0.10	0.07	0.10
WEE1-7	chr4	119010382	0.3	taatAGATGTTCTATTACTaAGG	0.04	0.03	0.04	0.06	0.05	0.06
WEE1-8	chr15	80845387	0.3	aTtaAGATGTTCTATTAaTaGGG	0.25	0.23	0.22	0.24	0.20	0.29
WEE1-9	chr13	87867165	0.2	cagctaATGTaaTATTACTCAGG	0.05	0.03	0.04	0.04	0.03	0.06
WEE1-10	chr4	107037600	0.2	aTCaAGATaTTCTATTACTgGGG	0.11	0.10	0.14	0.07	0.11	0.10
WEE1-11	chrX	136297269	0.2	tcttAtATtTTCTATTACTCAGG	0.24	0.22	0.26	0.29	0.22	0.18
WEE1-12	chr1	8500986	0.2	aTaGAGcaGTTaTATTACTCTGG	No PCR					
WEE1-13	chr11	71806684	0.2	GTaGAGATGTTCTcTaACTCAGG	0.67	0.71	0.66	0.69	0.69	0.70
WEE1-14	chr12	83599167	0.2	cTCttttatTTCTATTACTCTGG	0.00	0.00	0.00	0.00	0.00	0.00
WEE1-15	chr8	71394697	0.1	aatataATaTTCTATTACTCAGG	0.14	0.13	0.12	0.11	0.15	0.13
WEE1-16	chr8	86942319	0.1	caactataaTTCTATTACTCTGG	0.00	0.01	0.03	0.03	0.01	0.02
WEE1-17	chr1	61165563	0.1	GgCtActgtcTCTATTACTCAGG	0.14	0.10	0.11	0.08	0.12	0.11
WEE1-18	chr20	45730192	0.1	aTttAGATaTTCTATTACTCTGA	0.09	0.09	0.06	0.11	0.11	0.08
WEE1-19	chr8	134795474	0.1	cTttAagataTCTATTACTCTGG	0.11	0.08	0.10	0.12	0.09	0.08
WEE1-20	chr2	128238175	0.1	caaGAGAcGcTCTATcACTCAGG	0.26	0.26	0.24	0.25	0.27	0.24
WEE1-21	chr7	31677975	0.1	aaacAGcTGTTCTATTACTCAGG	0.09	0.09	0.15	0.12	0.12	0.13

(Continued)

(Continued)

EPHB4										
					Base editing frequency (%)					
					Untreated	(+) ABE7.10				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites		gX_{17}	gX_{18}	GX_{19}	gX_{20}	ggX_{20}
EphB4-01	chr7	100411300	10.3	GCAGAATATTCGGACAAACACGG	0.12	0.26	0.95	57.62	33.54	22.95
EphB4-02	chr5	81021421	8.0	GCAGAATATTaCaGACAAACtAGG	0.15	0.16	0.13	0.23	0.14	0.13
EphB4-03	chr15	47712791	4.1	GCAGAATATcaGGACAAACAATG	0.09	0.11	0.09	0.11	0.10	0.11
EphB4-04	chr11	123329727	2.4	aCAGAATATTCaGACAAtCACAG	0.09	0.06	0.11	0.10	0.10	0.10
EphB4-05	chr1	93047514	2.2	GCAGtA-AaTaaGACAAACAAGG	0.14	0.12	0.12	0.13	0.11	0.13
EphB4-06	chr15	51198151	1.9	aCAaAATATTCGGACAggCACGG	No PCR					
EphB4-07	chr18	29712895	1.8	GCAGAATtTaaGGACAAgCAAGG	0.33	0.29	0.29	0.30	0.35	0.29
EphB4-08	chr3	55670295	1.8	GCAGAATA--aGGACAAACATGG	0.12	0.12	0.09	0.15	0.19	5.82
EphB4-09	chr13	25919264	1.4	cagtAATATTCaGACtAAACATGG	0.50	0.49	0.49	0.42	0.44	0.40
EphB4-10	chr3	122057177	0.9	ttAatATATaaGGACAAACATGG	0.07	0.07	0.09	0.08	0.09	0.07
EphB4-11	chr1	26891812	0.8	aCAtAATATTCaaACAAACAGGG	0.07	0.07	0.07	0.08	0.07	0.05
EphB4-12	chr10	131074805	0.6	aatGAATAaagGGACAAACAGGG	0.52	0.57	0.57	0.54	0.53	0.58
EphB4-13	chr4	123425007	0.5	tCAG-ATATaaGGACAAACATGG	0.15	0.16	0.16	0.14	0.14	0.15
EphB4-14	chr5	60692862	0.4	GCtGAATATTgGGACAAtCATGG	0.15	0.14	0.12	0.14	0.13	0.14
EphB4-15	chr9	930844	0.4	GCAGttATAaaaaGACAAACAGGG	0.12	0.11	0.10	0.10	0.11	0.10
EphB4-16	chr9	29344386	0.4	aCAGAATATaaGaACAAAtATGG	0.06	0.06	0.06	0.08	0.05	0.06
EphB4-17	chr4	160752620	0.4	atAGAATAaTCaGACAAAaAAGG	0.20	0.16	0.18	0.20	0.17	0.18
EphB4-18	chr16	23706770	0.4	GCAG-ATAaagGGACAAACATGG	0.60	0.64	0.60	0.69	0.74	0.64
EphB4-19	chr11	117354211	0.3	GacagtgAaTaGGACAAACATGG	0.11	0.10	0.09	0.09	0.07	0.10
EphB4-20	chr14	90497703	0.3	tgAGAATATgCGGACAAACACAG	0.23	0.11	0.19	0.20	0.19	0.18
EphB4-21	chr2	68459706	0.3	atAGAATATTaGGACAAAtAGAG	No PCR					
EphB4-22	chr13	51951355	0.2	caAtAATAaaaGGACAAACATGG	0.04	0.04	0.06	0.04	0.05	0.03
EphB4-23	chr17	7777056	0.2	aCAGgATAaaaGGACAAAtATGG	0.05	0.05	0.06	0.06	0.08	0.08
EphB4-24	chr3	188431448	0.2	aCAGAtATAcgtaGACAAACAGAG	0.27	0.22	0.30	0.31	0.35	0.32
EphB4-25	chr12	11705824	0.2	atAGAATAagaaGACAAACATGG	0.06	0.05	0.09	0.06	0.05	0.06
EphB4-26	chr10	95697506	0.2	ttAGAATATTCctACAAACAAGG	0.13	0.10	0.10	0.10	0.10	0.11
EphB4-27	chrX	123665071	0.2	cCAtgtTATTCaGACAAACATGG	No PCR					
EphB4-28	chr16	75869847	0.2	GCAGAA-AaTaGGACAAtCACGG	0.29	0.26	0.33	0.27	0.31	0.31
EphB4-29	chr6	128784409	0.2	agAGAATATa-GGACAAACtTGG	0.55	0.40	0.37	0.48	0.37	0.40
EphB4-30	chr5	144297821	0.2	GCAGAAataTttGACAAACAAGG	0.10	0.05	0.05	0.08	0.07	0.07
EphB4-31	chr14	29760524	0.1	aaAGAATATa-GaACAAACAGGG	0.15	0.13	0.15	0.18	0.19	0.15
EphB4-32	chr1	36914336	0.1	aCAGAA-ATaCGaACAAACATGG	0.54	0.44	0.36	0.45	0.46	0.48
EphB4-33	chr21	26227853	0.1	aacaAATgcaaGGACAAACATGG	0.23	0.21	0.15	0.15	0.15	0.23
EphB4-34	chr13	63528393	0.1	aCActAT-TTaaGACAAACAGGG	0.15	0.14	0.14	0.13	0.14	0.14
EphB4-35	chr10	72516747	0.1	aCAGAActacgaGACAAACACGG	6.14	5.82	6.08	6.02	6.27	5.68
EphB4-36	chr7	20108292	0.1	GCAGA-TAccCaGACAAACAGGG	No PCR					
EphB4-37	chr1	89896856	0.1	taAGAATAT-aGGACAAACAGGA	0.22	0.16	0.21	0.19	0.17	0.17
EphB4-38	chr4	145533104	0.1	cCAGAATATTa-GACAAACATGG	0.12	0.16	0.14	0.12	0.14	0.13
EphB4-39	chr20	9784096	0.1	agAGAATATagGGACAAAtAAGG	0.11	0.08	0.09	0.11	0.09	0.11

HPRT-Exon6											
					Base editing frequency (%)						
						(+) ABE7.10					
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	gX17	gX18	GX19	gX20	ggX20
HPRT_E6-01	chrX	136276607	29.0	GTATAATCCAAAGATGGTCAAGG	0.11	16.94	29.23	62.07	60.39	58.46
HPRT_E6-02	chr12	7468915	21.3	GTATAcatCcAAGATGGcCAGGG	0.22	0.15	0.17	0.21	0.18	0.17
HPRT_E6-03	chr3	10057621	17.0	GTATtA-CCAAAGATGGTCtGGG	0.08	0.11	0.10	0.18	0.10	0.07
HPRT_E6-04	chr22	39586062	12.7	GTATAATC-AAAGATGGcCcTGG	0.38	0.45	0.42	0.40	0.43	0.39
HPRT_E6-05	chr7	13928792	10.5	GTATtAT-CAAAaATGGTCAAGG	0.15	0.22	0.14	0.14	0.15	0.15
HPRT_E6-06	chr1	225831201	10.2	aTATAA-CCAAAGATGtTCACAG	0.30	0.23	0.35	0.31	0.36	0.28
HPRT_E6-07	chr17	53575925	9.3	aTATAATCttAAGATGGTCAAGG	0.02	0.03	0.05	1.59	0.06	0.06
HPRT_E6-08	chr12	13406973	9.0	aTATAATCCtAAGATGtTCATGG	0.12	0.10	0.08	0.12	0.11	0.10
HPRT_E6-09	chr14	27255077	6.1	GTATtATCCAAAGATGacCAGAG	0.07	0.03	0.06	0.05	0.04	0.06
HPRT_E6-10	chr7	85914581	5.6	aTAcAATCCAAAGctTGGTCtGGG	0.24	0.19	0.15	0.14	0.18	0.12
HPRT_E6-11	chr3	9906633	4.7	aTATAttaCCAAAGATGGTCtGGG	0.15	0.22	0.25	0.17	0.15	0.16
HPRT_E6-12	chr8	125496169	4.5	GTAgAATCCAtAGATGGaCAGGG	0.26	0.22	0.26	0.31	0.26	0.26
HPRT_E6-13	chr17	10064482	4.3	GTATtA-CCAAAGATGGTCATGG	0.35	0.55	0.45	1.89	1.26	0.59
HPRT_E6-14	chr2	53374352	4.3	GTAcAATCaAAGATGaaCAAGG	0.10	0.11	0.13	0.11	0.08	0.09
HPRT_E6-15	chr4	102026801	4.1	aTATtATCCAAAGATGGagtTGG	0.07	0.05	0.09	0.06	0.04	0.08
HPRT_E6-16	chr6	19070067	3.8	aTgTAtTCCAAAGATGGTCAGGG	0.10	6.37	0.26	0.34	0.08	0.08
HPRT_E6-17	chr6	7761387	3.7	aagTAATCCAAAGATaGTCtTGG	0.08	0.17	0.10	0.09	0.10	0.12
HPRT_E6-18	chr12	59613203	3.6	GTATAAcCCAAAGAaaGTaATGG	0.26	0.31	0.37	0.31	0.37	0.39
HPRT_E6-19	chr16	20336811	3.6	aTATArATtCcAAGATGGTtcTGG	0.06	0.06	0.09	0.05	0.04	0.05
HPRT_E6-20	chr4	136809212	3.5	tTATAATCCAAAGAAGGCCAAGG	0.00	0.09	0.08	0.00	0.09	0.00
HPRT_E6-21	chr12	117915656	3.5	GTATctAcCCAAAGATGtTCATGG	0.07	0.06	0.11	0.08	0.09	0.08
HPRT_E6-22	chr14	45005861	3.4	GTATAAT--AAAGATGGTCAAGG	0.16	0.15	0.19	0.21	0.17	0.14
HPRT_E6-23	chr14	104033102	3.3	aTAGaTCCAAAGATGGaCATGG	0.33	0.26	0.27	0.22	0.25	0.28
HPRT_E6-24	chr11	93731924	3.1	cTATAATCtAAAaATGGTCAAGG	0.09	0.07	0.07	0.06	0.09	0.11
HPRT_E6-25	chr13	75257165	2.9	GTATAtTagAAAGATGGTCATGA	0.13	0.11	0.11	0.10	0.16	0.12
HPRT_E6-26	chr15	54911620	2.8	GgATAATCaAAAGATGGaCtTGG	0.11	0.08	0.05	0.07	0.12	0.09
HPRT_E6-27	chrX	6849511	2.6	aTATAATgCAAAGcTGGTCACGG	0.22	0.28	0.27	0.27	0.27	0.22
HPRT_E6-28	chr11	87526371	2.5	aTAcAATCCAAAGATatTCTtAGG	0.33	0.39	0.32	0.29	0.35	0.29
HPRT_E6-29	chr3	19652583	2.4	GTATAATCCAAAGgTGGccTAGG	0.23	0.21	0.16	0.24	0.20	0.19
HPRT_E6-30	chr17	45009128	2.4	aTATAATCCAtAGAaGGTCAAGA	0.50	0.46	0.50	0.49	0.51	0.45
HPRT_E6-31	chr8	9090663	2.3	GatatAatacAAGATGGaCAAGG	0.21	0.25	0.23	0.24	0.24	0.17
HPRT_E6-32	chr12	6810501	2.2	GgATAATCCAAAGATGGTgcAAG	0.24	0.27	2.20	0.29	0.25	0.28
HPRT_E6-33	chr6	118233456	1.5	aTATAAaCCAAAGATCcGTtAAGG	0.22	0.24	0.17	0.23	0.22	0.17
HPRT_E6-34	chr14	85667729	1.4	tctccTCCAAAGAaGGTCACGG	0.00	0.00	0.00	0.00	0.00	0.00
HPRT_E6-35	chrX	110551547	1.4	tTATAgTCCAAAGgaaGTCAGGG	0.20	0.16	0.24	0.22	0.23	0.23
HPRT_E6-36	chr8	18630339	1.2	tatTtATCCAAAGATGGTCtAGA	0.05	0.04	0.03	0.07	0.05	0.05
HPRT_E6-37	chr12	45338334	1.1	aTATAtaCCAAAGgTaGTCATGG	0.02	0.08	0.03	0.01	0.03	0.08
HPRT_E6-38	chr7	135056752	1.1	aggagATCCAAAGATGGTCAAGG	0.16	0.33	0.22	0.25	0.22	0.19
HPRT_E6-39	chr9	82090375	1.1	acATAgTCCAAAtATGGaCAGGG	0.04	0.06	0.04	0.04	0.04	0.04

(Continued)

HPRT_E6-40	chr12	103187212	1.0	GTATAcAattcAAGATGGTCAGGA	0.12	0.13	0.12	0.09	0.10	0.12
HPRT_E6-41	chr12	58996606	1.0	GaATAAgCCAAAGATGGTCAGTG	0.25	0.19	0.11	0.30	0.40	0.10
HPRT_E6-52	chr12	29153372	0.7	GTATttTCCAAcGATGGTCATGG	0.00	0.00	0.00	0.00	0.00	0.00
HPRT_E6-81	chr11	44024764	0.4	GTATAAgCCAAAGcaGGTCACGG	0.21	0.20	0.19	0.23	0.22	0.23

HPRT-Exon8

					Base editing frequency (%)					
						(+) ABE7.10				
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	gX_{17}	gX_{18}	GX_{19}	gX_{20}	ggX_{20}
HPRT_E8-1	chrX	133632686	2.4	GAAGTATTCATTATAGTCAAGGG	0.02	37.90	62.81	70.04	63.66	56.04
HPRT_E8-2	chr11	93382674	1.3	GAAGcATTCATTATAGTCAAAGG	0.20	0.20	0.25	5.77	0.56	0.29
HPRT_E8-3	chr21	34111536	0.6	GAAGaATTCATTATAGaCAATGG	0.16	0.13	0.19	0.16	0.15	0.12
HPRT_E8-4	chr11	100880874	0.6	aAgcTATTCATTATAGcaAATGG	0.04	0.05	0.05	0.06	0.04	0.06
HPRT_E8-5	chr22	29285654	0.6	GAAGTtaATTCATTATAGaCAATGG	0.10	0.12	0.12	0.08	0.11	0.13
HPRT_E8-6	chr4	15866630	0.5	aAAGTATTaATTATAGTCAAGGA	0.03	0.03	0.04	0.07	0.02	0.06
HPRT_E8-7	chr2	53543615	0.5	GAAGTAT-CcTTATAGTCAgAAG	0.16	0.11	0.12	0.15	0.16	0.15
HPRT_E8-8	chr3	149876617	0.4	GAAGTAT-CATTATAGTCttGGG	0.07	0.07	0.09	0.07	0.09	0.09
HPRT_E8-9	chr8	71429233	0.4	GcAGTATTCATaATAGTCAAAAG	0.02	0.02	0.02	0.03	0.04	0.01
HPRT_E8-10	chrX	146032194	0.4	aAAGTATTtcaTATAGccCAAAGG	0.01	0.01	0.01	0.01	0.01	0.02
HPRT_E8-11	chr5	132311070	0.4	GAAttTATTCAT-cATAGcCAAAGG	0.08	0.05	0.08	0.07	0.05	0.09
HPRT_E8-12	chr3	36349740	0.3	tAAGTAaTCATTATAGTCAgATG	0.10	0.06	0.10	0.09	0.08	0.09
HPRT_E8-13	chr3	181655989	0.3	atAGTATTCATTATAGTaAcAGG	0.07	0.04	0.03	0.03	0.04	0.03
HPRT_E8-14	chr11	116029576	0.3	aAAaTATTcCATTATAGTCttACG	0.08	0.11	0.10	0.14	0.11	0.10
HPRT_E8-15	chr10	131905650	0.3	GAAcTATTCATTATcaTCAACGG	0.10	0.10	0.07	0.07	0.10	0.10
HPRT_E8-16	chr9	89312308	0.2	Gttt-GTATTCATTATAGTCAAGAG	0.02	0.05	0.05	0.02	0.02	0.02
HPRT_E8-17	chr4	178543184	0.2	atAGTATTCATTATAGTaAcAGG	0.10	0.09	0.11	0.11	0.10	0.10
HPRT_E8-18	chr12	72733535	0.2	agAGTATcCAccATAGTCAAGGG	0.05	0.04	0.45	0.08	0.04	0.04
HPRT_E8-19	chr1	45916291	0.1	GAAGTAcatcTTATAaTCAAATGG	0.05	0.05	0.03	0.06	0.03	0.06
HPRT_E8-20	chr4	153599265	0.1	aATGTATTCATTATAGgCAAGAG	0.15	0.14	0.18	0.17	0.17	0.17

(With Daesik Kim in Institute for Basic Science)

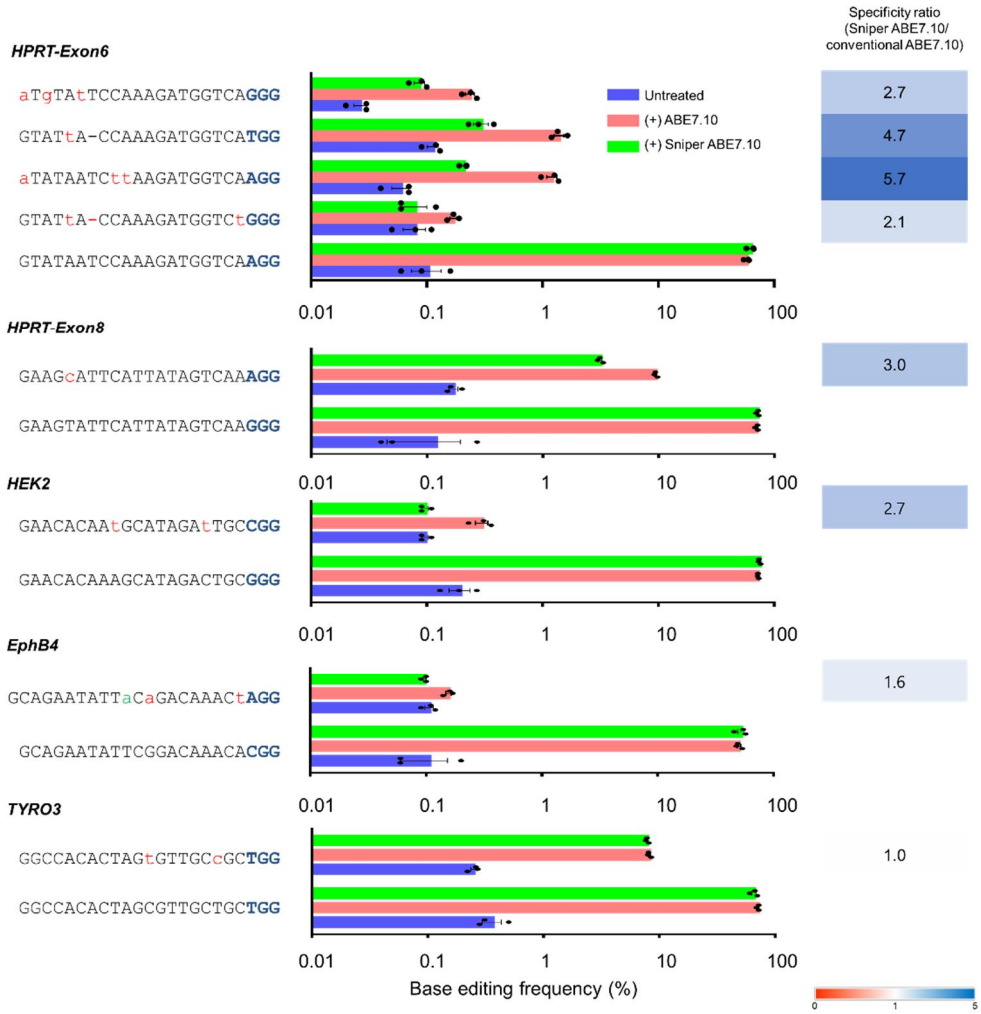


Figure 17. Increasing specificity using Sniper ABE7.10. On- and off-target activity and specificity ratio when using ABE7.10 and Sniper ABE7.10 at the *HPRT*-exon6, *HPRT*-exon8, *HEK2, EphB4 and TYRO3* sites. The red and blue characters indicate the mismatched nucleotides and PAM sequences, respectively. The heatmap shows relative specificities of Sniper ABE7.10 compared to that of plasmid DNA encoding ABE7.10. The specificity ratio was calculated using the

following formula: specificity of Sniper ABE7.10 (on-target frequency/off-target frequency)/ specificity of conventional ABE7.10 (on-target frequency/off-target frequency). Means ± s.e.m. were from three independent experiments.

Table 8. Mutation frequencies of ABE7.10 and Sniper ABE7.10 at on-target and off-target sites captured by Digenome-seq

HEK2							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
HEK2_001	chr19	35505485	20.1	GAACAC-AAGCAcAGACTGaAGG	0.17	0.17	0.16
HEK2_002	chr15	93557679	15.1	GAACACA-tGCATAGACTGCTAG	0.55	0.55	0.59
HEK2_003	chr4	53536209	14.7	GAAtACtAAGCATAGACTcCAGG	0.02	0.05	0.07
HEK2_004	chr4	90522183	14.3	GAACACAAtGCATAGAtTGCCGG	0.16	0.33	0.13
HEK2_005	chr1	167742859	13.4	aAACACAgAGCAcAGACTGCTGA	0.16	0.22	0.22
HEK2_006	chr10	43159340	13.4	aAACACAAAGaCATAGACcaCTGG	0.22	0.18	0.17
HEK2_007	chr5	87240613	8.9	GAACACAAAGCATAGACTGCGGG	0.31	72.76	75.92
HEK2_008	chr9	290167	6.6	aAACAtAAAGaATAGACTGCAAG	0.09	0.10	0.11
HEK2_009	chr3	28560566	5.8	GAACACAAAGtATAGAaTGCTAG	0.11	0.10	0.09
HEK2_010	chr1	175310138	5.1	GgACACAAAGCtTAGACTcCAGG	0.39	0.42	0.34
HEK2_011	chr13	87616351	4.9	GAAgAtAAAGCATAGACTctAGG	0.12	0.13	0.11
HEK2_012	chr10	87125365	4.6	aAACAtAAAGCATAGACTGCaAAG	0.24	0.23	0.22
HEK2_013	chr18	32279782	3.9	aAAtACAAAGCATAGACTaatATG	0.20	0.24	0.20
HEK2_014	chr2	192248363	3.5	GAACACAtA-CATAGACaGCTGG	0.02	0.03	0.00
HEK2_015	chr5	160202385	3.0	GAAaACAAAGCAaAGAaaGCAGG	0.17	0.21	0.22
HEK2_016	chr15	63236612	2.9	aAcacaAtAGCATAGACTGgaCTG	No PCR		
HEK2_017	chr20	23101377	2.7	atACACAgAGCAaAGACTGCAGG	0.20	0.22	0.18
HEK2_018	chrX	48317375	2.5	aAACAtAAAGCcTAGACTGaCGG	0.32	0.30	0.31
HEK2_019	chr19	49972515	2.2	aAAaACAAAGCgTAGACTGtGGG	0.37	0.36	0.43
HEK2_020	chr12	81785997	2.1	GgACAaAAAGCATAGACTcaAGG	0.15	0.13	0.12
HEK2_021	chr1	40461424	2.1	atACACACAAGCAcAGACTGCAGG	0.50	0.42	0.47
HEK2_022	chr15	92968604	2.1	GAACACAgcaCATAaACTGCAGG	0.19	0.20	0.24
HEK2_023	chr5	142142325	2.0	aAACACtAA-CATAGACTGCAGG	6.76	6.74	7.27
HEK2_024	chr5	131174461	1.7	aAAtACAAtGCATAGACTGCTAG	0.10	0.11	0.11
HEK2_025	chr1	36097072	1.4	GtAaACAAAGCATAGACTGaGGG	0.22	0.24	0.23
HEK2_026	chrX	132724931	1.3	GAACA-tAAtCAcAGACTGCTGG	0.23	0.19	0.22
HEK2_027	chr6	139353017	1.3	ccAaACAAAaCATAGACTGCTGG	0.40	0.39	0.41
HEK2_028	chrX	36949816	1.3	GAAaACAAAaCATAGAgTGCTGG	0.14	0.16	0.15
HEK2_029	chr17	2670110	1.2	aAtCAaAtAGCATAGACTGCATG	0.22	0.20	0.23
HEK2_030	chr13	91254877	1.2	GAACACAAAcatagtACTGaAGG	0.18	0.20	0.16
HEK2_031	chrX	138877515	1.2	aAACAgAAAGCATgGACTGCGGA	0.10	0.16	0.13
HEK2_032	chr13	113428466	1.1	cAAtACAAAGgATAGACTGCAGG	0.08	0.12	0.10
HEK2_033	chr15	41796089	1.1	agACACAcAGCAcAGACTGCAGG	0.18	0.21	0.20
HEK2_034	chr11	128508576	1.0	GAAttCAAAGCATAGAtTGCAGG	0.13	0.11	0.10
HEK2_042	chr3	33744669	0.7	GAAaACAAAcCAcAGACTGgGGG	0.20	0.21	0.24
HEK2_044	chr16	64198277	0.7	GAAtACAtAaCATAGACTGgGGG	0.00	0.00	0.00
HEK2_050	chr20	97640	0.5	GAAttCAAAGCATAGAtTGCAGG	0.12	0.13	0.13
HEK2_052	chr3	10835885	0.5	aAACACAtAGCAcAGcCTGCAGG	0.13	0.15	0.16
HEK2_065	chr18	56307002	0.4	aAgaACAAAaCATAGACTGCAGG	0.23	0.26	0.22
HEK2_079	chr5	126385455	0.3	ccACACcAAGCATAGACTtCTGG	0.12	0.09	0.13
HEK2_082	chr12	131431445	0.3	GgAgAgAgAGCATAGACTGCTGG	0.21	0.22	0.21
HEK2_092	chr1	237916998	0.2	tAACACAAAGacTAaACTGCAGG	0.10	0.10	0.11
HEK2_093	chr1	215323272	0.2	tAcCACgAAGCATAGACTGtAGG	0.12	0.10	0.09
HEK2_099	chr13	45266353	0.2	GgACAtAAAtCATAaACTGCTGG	0.12	0.14	0.10
HEK2_101	chr1	77190607	0.2	tcACACAAAcCATAGACTGaGGG	0.19	0.18	0.17
HEK2_105	chr19	5068807	0.2	tAtCAgAAAGCAcAGACTGCGGG	0.16	0.14	0.14
HEK2_106	chr12	54405073	0.2	GAACAacAAGgATAaACTGCCGG	0.27	0.29	0.30
HEK2_110	chr20	17801623	0.2	GAAtgCcAAGCATAaACTGCAGG	0.00	0.00	0.00
HEK2_119	chr19	32225924	0.2	GgACACAgAGtATAGACTGaGGG	0.10	0.12	0.13
HEK2_121	chr13	58440671	0.1	GAAttCAAAGCATgGACTGCAGG	0.09	0.11	0.10
HEK2_139	chr7	113235709	0.1	tAAaACAAAGaAcAGACTGCTGG	0.25	0.31	0.28
HEK2_145	chr14	88566211	0.1	GtAgAaAAAGtATAGACTGCAGG	0.15	0.14	0.15

(Continued)

RNF2							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
RNF2-1	chr1	210858376	7.3	tcaccaTTAGTCATTACCTGCTG	0.03	0.05	0.05
RNF2-2	chr1	185056773	4.4	GTCATCTTAGTCATTACCTGAGG	0.22	32.79	16.24
RNF2-3	chr17	53928598	4.3	AGTCATCTTAGTCATTAC-TGAGG	0.44	0.45	0.51
RNF2-4	chr6	143212078	4.1	GTaATaTTAGTCATTACCgGTGG	0.18	0.17	0.19
RNF2-5	chr17	17072075	3.6	GTCATCcTAGTCATTtaCTGGGG	0.98	0.94	0.89
RNF2-6	chr12	131619804	2.2	aTCAcCTTAGcCATTACCaGGGG	0.35	0.33	0.42
RNF2-7	chr9	102127746	1.4	GTCA-CTTAGTCATTgCCTGTGG	0.36	0.37	0.33
RNF2-8	chr15	92315580	1.0	GTCAcaTTAGcCATTACCTGTGA	0.32	0.35	0.27
RNF2-15	chr2	51894204	0.4	aTCATCTTcaTCATTACaTGAGG	0.29	0.27	0.26
RNF2-19	chr4	125521739	0.4	tTaATCTTAGTCATTACtTtTGG	0.22	0.23	0.24
RNF2-20	chr2	102668931	0.3	aTCATCaTcGTCATTAtCTGGGG	1.39	1.23	1.22
RNF2-24	chr1	111785389	0.3	tcCATCTcAcTCATTACCTGAGG	0.11	0.18	0.18
RNF2-27	chr6	5869239	0.3	GcCAcCTcAGTCATTAgCTGGGG	0.40	0.48	0.46
RNF2-41	chr18	62061226	0.2	aTtATtTTAGTCATTACCTtTGG	0.02	0.02	0.04

TYRO3							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
TYRO3-1	chr15	76553332	4.3	GGCCACACTAGtGTTGCcGCTGG	0.64	3.04	3.12
TYRO3-2	chr15	41857272	3.4	GGCCACACTAGCGTTGCTGCTGG	1.17	72.66	68.95
TYRO3-3	chr20	45312757	2.0	GGCCACACcAGCcTTGCTGtCGG	11.24	11.54	11.48
TYRO3-4	chr5	173133468	0.8	GGCCAC--TAGCGTTGCTcCAGG	0.36	0.36	0.32
TYRO3-5	chr22	50686658	0.6	GGCCACACTgaCcTTGCTGCTGG	1.02	1.07	1.02
TYRO3-6	chr9	140109027	0.4	ttCCACAC-AGCtTTGCTGCTGG	0.68	0.63	0.61
TYRO3-7	chr17	17207899	0.3	GGCCAC-CTAGgGTTGCTGCTGG	0.54	0.44	0.76
TYRO3-8	chr11	99666699	0.2	ccaCcCtCTAGCaTTGCTGCTGG	0.00	0.00	0.00
TYRO3-9	chr11	27495183	0.2	cctaACACcAaCGTTGCTGCTGG	0.98	0.85	0.95
TYRO3-10	chr1	38684664	0.2	ctgagCtCTAGCaTTGCTGCTGG	0.16	0.18	0.07
TYRO3-11	chr3	22936710	0.1	GGCatataTAGCaTTGCTGCTGG	0.32	0.22	0.28
TYRO3-12	chr6	150629124	0.1	GaaCACACTAaCaTTGCTGtGGG	0.70	0.88	0.82

WEE1							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
WEE1-1	chr1	234241713	11.1	tTaGAGATGTTCTATTAtTCCGG	0.80	0.73	0.76
WEE1-2	chr11	9610086	2.0	GTCGAGATGTTCTATTACTCTGG	0.35	58.59	53.38
WEE1-3	chr4	18855558	1.9	GagGAGATGcTCTATTACTCCGG	2.38	2.32	2.08
WEE1-4	chrX	42756785	0.7	aTCaAGATGaaCTATTACTCTGG	0.21	0.24	0.22
WEE1-5	chr1	215805743	0.4	GTgGAGATGTTaTgTTACTCTGG	0.18	0.17	0.17
WEE1-6	chr5	57863350	0.4	GaCataATGTTCTATTACTCAAG	0.76	0.96	0.88
WEE1-7	chr4	119010382	0.3	taatAGATGTTCTATTACTaAGG	0.12	0.12	0.11
WEE1-8	chr15	80845387	0.3	aTtaAGATGTTCTATTAaTaGGG	1.05	0.92	1.02
WEE1-9	chr13	87867165	0.2	cagctaATGTaaTATTACTCAGG	0.59	0.61	0.58
WEE1-10	chr4	107037600	0.2	aTCaAGATaTTCTATTACTgGGG	0.24	0.28	0.23
WEE1-11	chrX	136297269	0.2	tcttAtATtTTCTATTACTCAGG	0.23	0.21	0.13
WEE1-12	chr1	8500986	0.2	aTaGAGcaGTTaTATTACTCTGG	No PCR		
WEE1-13	chr11	71806684	0.2	GTaGAGATGTTCTcTaACTCAGG	1.45	1.50	1.65
WEE1-14	chr12	83599167	0.2	cTCtttttatTTCTATTACTCTGG	0.00	0.00	0.00
WEE1-15	chr8	71394697	0.1	aatataATaTTCTATTACTCAGG	0.46	0.41	0.56
WEE1-16	chr8	86942319	0.1	caactataaTTCTATTACTCTGG	0.00	0.00	0.00
WEE1-17	chr1	61165563	0.1	GgCtActgtcTCTATTACTCAGG	0.18	0.17	0.13
WEE1-18	chr20	45730192	0.1	aTttAGATaTTCTATTACTCTGA	0.23	0.28	0.24
WEE1-19	chr8	134795474	0.1	cTttAagataTCTATTACTCTGG	0.32	0.45	0.32
WEE1-20	chr2	128238175	0.1	caaGAGAcGcTCTATcACTCAGG	0.81	0.97	1.13
WEE1-21	chr7	31677975	0.1	aaacAGcTGTTCTATTACTCAGG	0.21	0.24	0.25

(Continued)

EphB4							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
EphB4-01	chr7	100411300	10.3	GCAGAATATTCGGACAAACACGG	0.13	52.42	54.66
EphB4-02	chr5	81021421	8.0	GCAGAATATTaCaGACAAACtAGG	0.15	0.19	0.13
EphB4-03	chr15	47712791	4.1	GCAGAATATcaGGACAAACAATG	0.16	0.16	0.16
EphB4-04	chr11	123329727	2.4	aCAGAATATTCaGACAAtCACAG	0.26	0.27	0.21
EphB4-05	chr1	93047514	2.2	GCAGtA-AaTaaGACAAACAAGG	0.12	0.14	0.12
EphB4-06	chr15	51198151	1.9	aCAaAATATTCGGACAggCACGG	No PCR		
EphB4-07	chr18	29712895	1.8	GCAGAATtTaaGGACAAgCAAGG	0.13	0.14	0.14
EphB4-08	chr3	55670295	1.8	GCAGAATA--aGGACAAACATGG	No PCR		
EphB4-09	chr13	25919264	1.4	cagtAATATTCaGACtAAACATGG	0.27	0.24	0.26
EphB4-10	chr3	122057177	0.9	ttAatATATaaGGACAAACATGG	0.07	0.08	0.07
EphB4-11	chr1	26891812	0.8	aCAtAATATTCaaACAAACAGGG	0.14	0.13	0.14
EphB4-12	chr10	131074805	0.6	aatGAATAaagGGACAAACAGGG	0.18	0.19	0.22
EphB4-13	chr4	123425007	0.5	tCAG-ATATaaGGACAAACATGG	0.18	0.19	0.17
EphB4-14	chr5	60692862	0.4	GCtGAATATTgGGACAAtCATGG	0.12	0.10	0.13
EphB4-15	chr9	930844	0.4	GCAGttATAaaaaGACAAACAGGG	0.02	0.02	0.03
EphB4-16	chr9	29344386	0.4	aCAGAATATaaGaACAAAtATGG	0.05	0.07	0.08
EphB4-17	chr4	160752620	0.4	atAGAATAaTCaGACAAAaAAGG	0.09	0.08	0.08
EphB4-18	chr16	23706770	0.4	GCAG-ATAaagGGACAAACATGG	0.39	0.37	0.35
EphB4-19	chr11	117354211	0.3	GacagtgAaTaGGACAAACATGG	0.06	0.07	0.10
EphB4-20	chr14	90497703	0.3	tgAGAATATgCGGACAAACACAG	0.11	0.13	0.12
EphB4-21	chr2	68459706	0.3	atAGAATATTaGGACAAAtAGAG	0.00	0.16	0.15
EphB4-22	chr13	51951355	0.2	caAtAATAaaaGGACAAACATGG	0.08	0.11	0.08
EphB4-23	chr17	7777056	0.2	aCAGgATAaaaGGACAAAtATGG	0.06	0.10	0.07
EphB4-24	chr3	188431448	0.2	aCAGAtATAcgtaGACAAACAGAG	0.12	0.14	0.14
EphB4-25	chr12	11705824	0.2	atAGAATAagaaGACAAACATGG	0.30	0.38	0.35
EphB4-26	chr10	95697506	0.2	ttAGAATATTCctACAAACAAGG	0.07	0.09	0.08
EphB4-27	chrX	123665071	0.2	cCAtgtTATTCaGACAAACATGG	No PCR		
EphB4-28	chr16	75869847	0.2	GCAGAA-AaTaGGACAAtCACGG	0.17	0.14	0.14
EphB4-29	chr6	128784409	0.2	agAGAATATa-GGACAAACtTGG	0.22	0.22	0.21
EphB4-30	chr5	144297821	0.2	GCAGAAataTttGACAAACAAGG	0.13	0.13	0.09
EphB4-31	chr14	29760524	0.1	aaAGAATATa-GaACAAACAGGG	0.24	0.20	0.22
EphB4-32	chr1	36914336	0.1	aCAGAA-ATaCGaACAAACATGG	0.21	0.16	0.23
EphB4-33	chr21	26227853	0.1	aacaAATgcaaGGACAAACATGG	0.17	0.18	0.12
EphB4-34	chr13	63528393	0.1	aCActAT-TTaaGACAAACAGGG	0.17	0.15	0.13
EphB4-35	chr10	72516747	0.1	aCAGAActacgaGACAAACACGG	2.32	2.36	2.33
EphB4-36	chr7	20108292	0.1	GCAGA-TAccCaGACAAACAGGG	No PCR		
EphB4-37	chr1	89896856	0.1	taAGAATAT-aGGACAAACAGGA	0.13	0.12	0.12
EphB4-38	chr4	145533104	0.1	cCAGAATATTa-GACAAACATGG	0.22	0.25	0.24
EphB4-39	chr20	9784096	0.1	agAGAATATagGGACAAAtAAGG	0.19	0.22	0.22

(Continued)

HPRT-Exon6							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
HPRT_E6-01	chrX	133627607	29.0	GTATAATCCAAAGATGGTCAAGG	0.73	58.63	63.33
HPRT_E6-02	chr12	7468915	21.3	GTATAcatCcAAGATGGcCAGGG	0.14	0.17	0.15
HPRT_E6-03	chr3	10057621	17.0	GTATtA-CCAAAGATGGTCtGGG	0.22	0.34	0.28
HPRT_E6-04	chr22	39586062	12.7	GTATAATC-AAAGATGGcCcTGG	0.34	0.36	0.35
HPRT_E6-05	chr7	13928792	10.5	GTATtAT-CAAAaATGGTCAAGG	0.12	0.12	0.12
HPRT_E6-06	chr1	225831201	10.2	aTATAA-CCAAAGATGtTCACAG	0.26	0.29	0.23
HPRT_E6-07	chr17	53575925	9.3	aTATAATCttAAGATGGTCAAGG	0.12	1.45	0.19
HPRT_E6-08	chr12	13406973	9.0	aTATAATCCtAAGATGtTCATGG	0.12	0.11	0.10
HPRT_E6-09	chr14	27255077	6.1	GTATtATCCAAAGATGacCAGAG	0.12	0.12	0.15
HPRT_E6-10	chr7	85914581	5.6	aTAcAATCCAAAGtTGGTCtGGG	0.09	0.11	0.11
HPRT_E6-11	chr3	9906633	4.7	aTATAttaCCAAAGATGGTCtGGG	0.26	0.32	0.24
HPRT_E6-12	chr8	125496169	4.5	GTAgAATCCAtAGATGGaCAGGG	0.18	0.18	0.16
HPRT_E6-13	chr17	10064482	4.3	GTATtA-CCAAAGATGGTCATGG	0.12	1.19	0.23
HPRT_E6-14	chr2	53374352	4.3	GTAcAATCaAAAGATGaaCAAGG	0.14	0.12	0.16
HPRT_E6-15	chr4	102026801	4.1	aTATtATCCAAAGATGGagtTGG	0.04	0.03	0.04
HPRT_E6-16	chr6	19070067	3.8	aTgTAtTCCAAAGATGGTCAGGG	0.07	0.19	0.07
HPRT_E6-17	chr6	7761387	3.7	aagTAATCCAAAGATaGTCtTGG	0.12	0.12	0.12
HPRT_E6-18	chr12	59613203	3.6	GTATAAcCCAAAGAaaGTaATGG	0.25	0.26	0.20
HPRT_E6-19	chr16	20336811	3.6	aTATAtATtCcAAGATGGTtcTGG	0.07	0.07	0.08
HPRT_E6-20	chr4	136809212	3.5	tTATAATCCAAAGAaGGcCAAGG	0.14	0.10	0.13
HPRT_E6-21	chr12	117915656	3.5	GTATctAcCCAAAGATGtTCATGG	0.07	0.08	0.07
HPRT_E6-22	chr14	45005861	3.4	GTATAAT--AAAGATGGTCAAGG	0.12	0.15	0.12
HPRT_E6-23	chr14	104033102	3.3	aTAgAtTCCAAAGATGGaCATGG	0.08	0.08	0.07
HPRT_E6-24	chr11	93731924	3.1	cTATAATCtAAAaATGGTCAAGG	0.11	0.12	0.12
HPRT_E6-25	chr13	75257165	2.9	GTATAtTagAAAGATGGTCATGA	0.09	0.08	0.09
HPRT_E6-26	chr15	54911620	2.8	GgATAATCaAAAGATGGaCtTGG	0.09	0.07	0.07
HPRT_E6-27	chrX	6849511	2.6	aTATAATgCAAAGcTGGTCACGG	0.46	0.45	0.45
HPRT_E6-28	chr11	87526371	2.5	aTAcAATCCAAAGATatTCtAGG	0.13	0.13	0.14
HPRT_E6-29	chr3	19652583	2.4	GTATAATCCAAAGgTGGcCtAGG	0.28	0.27	0.29
HPRT_E6-30	chr17	45009128	2.4	aTATAATCCAtAGAaGGTCAAGA	0.27	0.30	0.30
HPRT_E6-31	chr8	9090663	2.3	GatatAatacAAGATGGaCAAGG	0.15	0.18	0.12
HPRT_E6-32	chr12	6810501	2.2	GgATAATCCAAAGATGGTgcAAG	0.23	0.24	0.20
HPRT_E6-33	chr6	118233456	1.5	aTATAAaCCAAAGATcGTtAAGG	0.22	0.24	0.27
HPRT_E6-34	chr14	85667729	1.4	tctcccTCCAAAGAaGGTCACGG	0.00	0.00	0.00
HPRT_E6-35	chrX	110551547	1.4	tTATAgTCCAAAGgaaGTCAGGG	0.10	0.09	0.09
HPRT_E6-36	chr8	18630339	1.2	tatTtATCCAAAGATGGTCtAGA	0.09	0.11	0.08
HPRT_E6-37	chr12	45338334	1.1	aTATAtaCCAAAGgTaGTCATGG	0.02	0.06	0.09
HPRT_E6-38	chr7	135056752	1.1	aggagATCCAAAGATGGTCAAGG	0.21	0.22	0.21
HPRT_E6-39	chr9	82090375	1.1	acATAgTCCAAAtATGGaCAGGG	0.04	0.04	0.02
HPRT_E6-40	chr12	103187212	1.0	GTATAtAattcAAGATGGTCAGGA	0.23	0.23	0.28
HPRT_E6-41	chr12	58996606	1.0	GaATAAgCCAAAGATGGTCAGTG	0.06	0.13	0.16
HPRT_E6-52	chr12	29153372	0.7	GTATttTCCAAcGATGGTCATGG	0.00	0.00	0.00
HPRT_E6-81	chr11	44024764	0.4	GTATAAgCCAAAGcaGGTCACGG	0.13	0.18	0.15

HPRT-Exon8							
					Base editing frequency (%)		
	Chr	Position	DNA Cleavage Score	DNA seq at a cleavage sites	Untreated	(+) ABE7.10	(+) Sniper ABE7.10
HPRT_E8-1	chrX	133632686	2.4	GAAGTATTCATTATAGTCAAGGG	0.03	70.09	71.15
HPRT_E8-2	chr11	93382674	1.3	GAAGcATTCATTATAGTCAAAGG	0.11	10.33	3.53
HPRT_E8-3	chr21	34111536	0.6	GAAGaATTCATTATAGaCAATGG	0.17	0.17	0.19
HPRT_E8-4	chr11	100880874	0.6	aAgcTATTCATTATAGcaAATGG	0.05	0.03	0.06
HPRT_E8-5	chr22	29285654	0.6	GAAGTtaATTCATTATAGaCAATGG	0.08	0.08	0.10
HPRT_E8-6	chr4	15866630	0.5	aAAGTATTaATTATAGTCAAGGA	0.04	0.10	0.09
HPRT_E8-7	chr2	53543615	0.5	GAAGTAT-CcTTATAGTCAgAAG	0.16	0.16	0.14
HPRT_E8-8	chr3	149876617	0.4	GAAGTAT-CATTATAGTCttGGG	No PCR		
HPRT_E8-9	chr8	71429233	0.4	GcAGTATTCATaATAGTCAAAAG	0.07	0.06	0.04
HPRT_E8-10	chrX	146032194	0.4	aAAGTATTtcaTATAGcCAAAGG	0.11	0.11	0.12
HPRT_E8-11	chr5	132311070	0.4	GAAtTATTCATcATAGcCAAAGG	0.05	0.07	0.05
HPRT_E8-12	chr3	36349740	0.3	tAAGTAaTCATTATAGTCAgATG	0.14	0.15	0.13
HPRT_E8-13	chr3	181655989	0.3	atAGTATTCATTATAGTaAcAGG	0.05	0.05	0.03
HPRT_E8-14	chr11	116029576	0.3	aAAaTATcCATTATAGTCttACG	0.26	0.25	0.25
HPRT_E8-15	chr10	131905650	0.3	GAAcTATTCATTATtaTCAACGG	0.04	0.05	0.06
HPRT_E8-16	chr9	89312308	0.2	GttGTATTCATTATAGTCAAGAG	0.11	0.11	0.10
HPRT_E8-17	chr4	178543184	0.2	atAGTATTCATTATAGTaAcAGG	0.09	0.09	0.08
HPRT_E8-18	chr12	72733535	0.2	agAGTATcCAccATAGTCAAGGG	0.04	0.15	0.05
HPRT_E8-19	chr1	45916291	0.1	GAAGTAcatcTTATAaTCAATGG	0.05	0.04	0.07
HPRT_E8-20	chr4	153599265	0.1	aAtGTATTCATTATAGgCAAGAG	0.08	0.05	0.05

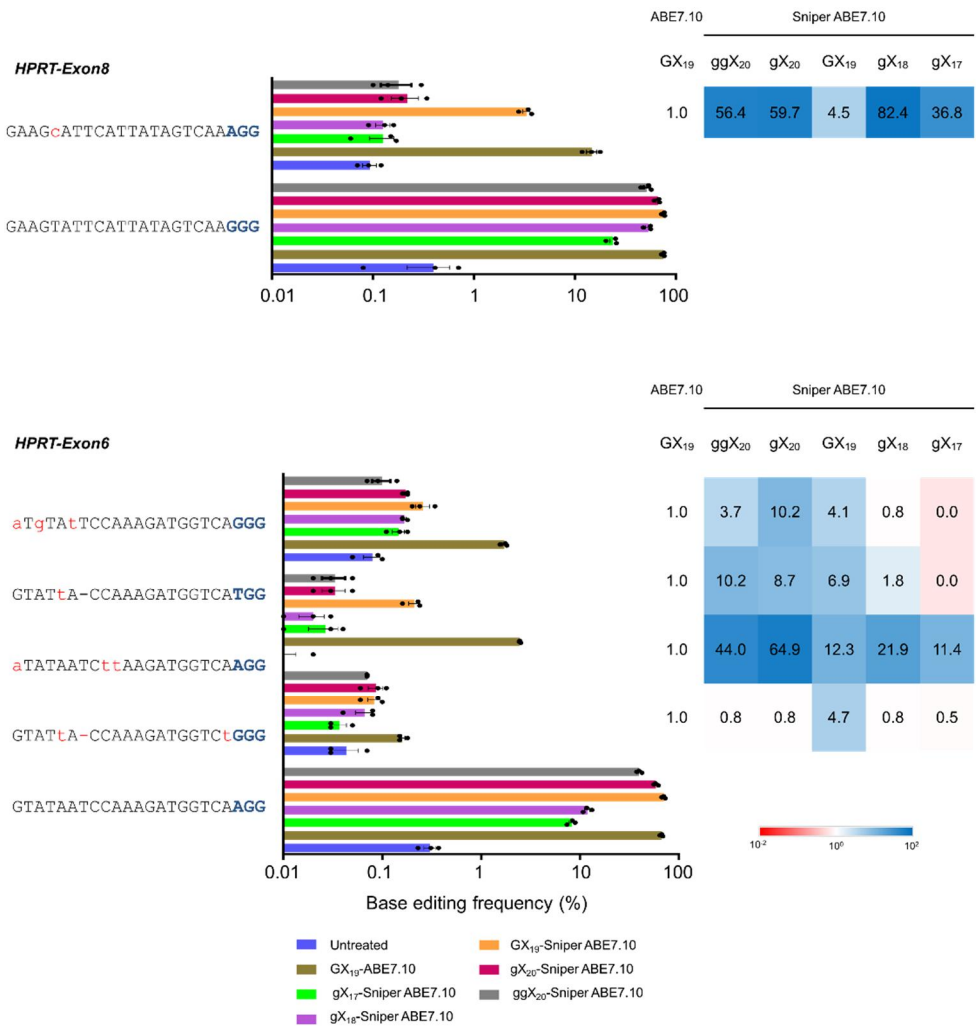


Figure 18. Increasing specificity of ABE7.10 by combining Sniper ABE7.10 with modified sgRNAs. On- and off-target activity and specificity ratio using ABE7.10 and Sniper ABE7.10 with modified sgRNAs at the *HPRT*-exon6 and *HPRT*-exon8 sites. The red and blue characters indicate the mismatched nucleotides and PAM sequences, respectively. The heatmap shows relative specificities of Sniper-ABE7.10 with modified sgRNAs compared to that of ABE7.10 and GX_{19} sgRNAs. The specificity ratio was calculated using the

following formula: specificity of ABE7.10 (on-target frequency/off-target frequency)/ specificity of Sniper ABE7.10 with modified sgRNAs. (on-target frequency/off-target frequency). Means ± s.e.m. were from three independent experiments.

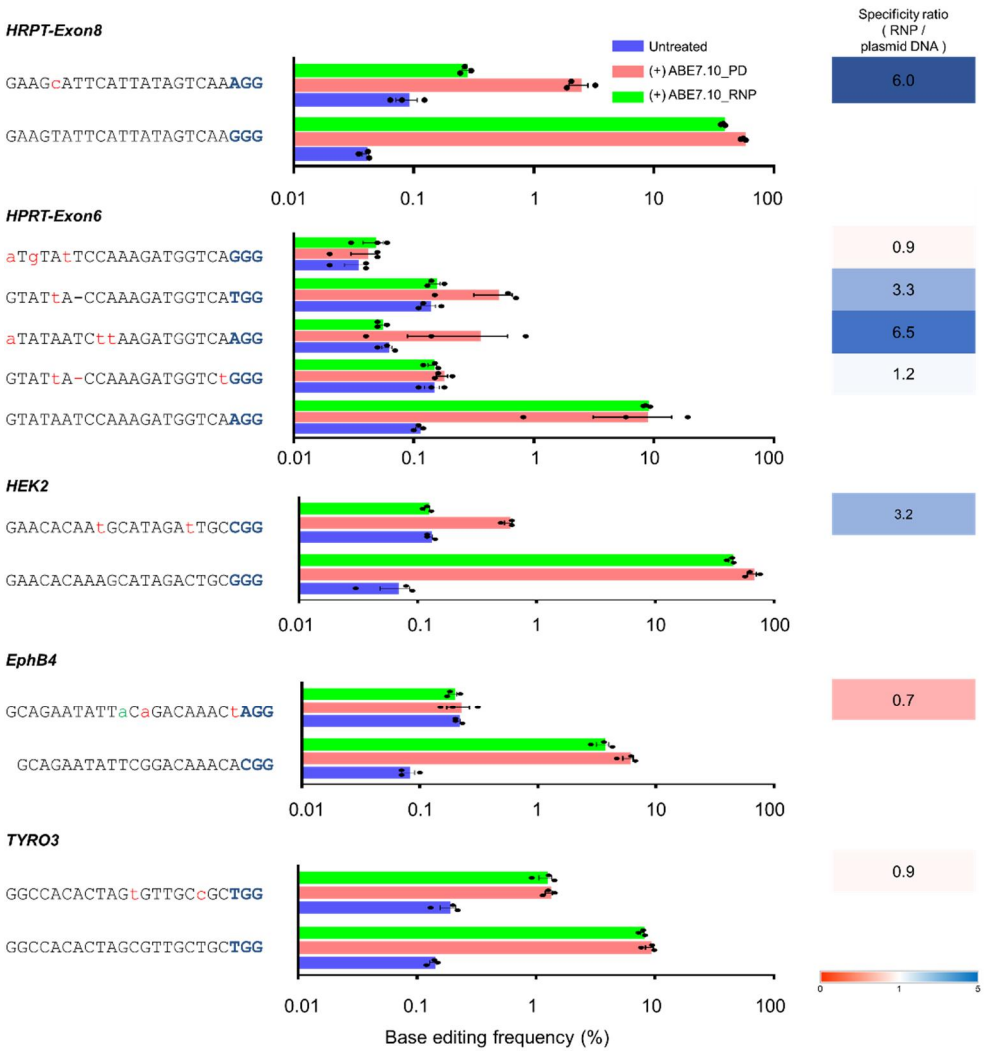


Figure 19. Specificity of ABE7.10 depending on plasmid delivery or RNP delivery. On- and off-target activity and specificity ratio when delivering plasmid or RNP at the *HPRT*-exon8, *HPRT*-exon6, *HEK2, EphB4 and TYRO3* sites. The red and blue characters indicate the mismatched nucleotides and PAM sequences, respectively. The heatmap shows relative specificities of ABE7.10 RNP delivery compared to that of plasmid DNA delivery encoding ABE7.10. The specificity ratio was calculated using the following formula: specificity of RNP delivery (on-target

frequency/off-target frequency)/ specificity of plasmid delivery (on-target frequency/off-target frequency). Means ± s.e.m. were from three independent experiments.

Ⅳ. Discussion

If the existing CRISPR/Cas9 system has been used primarily for genetic knockout studies, the CRISPR RNA-guide base editor has been recently reported to be able to introduce more sophisticated and highly efficient point mutations (Gaudelli et al., 2017; Hess et al., 2016; Jiang et al., 2018; Komor et al., 2016; Li et al., 2018b; Ma et al., 2016b; Nishida et al., 2016). ABEs developed by directed evolution methods following several CBE types have been reported and the scope of studying human diseases has been broadened.

In order to increase the efficiency of the base editor system to be more widely used, there have also been studies such as using an ancestral sequence of deaminase or giving a selection pressure (Koblan et al., 2018; Zafra et al., 2018). In addition, as in the previous reports (Koblan et al., 2018; Staahl et al., 2017; Zafra et al., 2018), this study confirms that it is also possible to optimize codon usage for the organism in which the base editor operates, for increasing protein expression or to change the type and number of NLSs, especially in cell lines expressing base editor constructs weakly.

I also compared the activity of the lower version of ABEs. It was confirmed that the activity ranking of varies according to the evolved version of TadA in HeLa and HEK293T cells. The activity of ABE7.10 was always best and the activity of ABE6.3 or ABE7.9 varied depending on the target site and cell type. I found that codon usage is important for other versions as well as ABE7.10, which is known to be the most active and widely used. Efficiency of ABE variants may

vary when used with various mismatched sgRNAs, so it may be helpful to use different versions of ABEs than only ABE7.10 if someone wants to reduce the efficiency of the off-target versus on-target at some target sites.

Newly known ABEs had to be proven to be genome-wide specific in order to be more widely used. I could confirm that mismatched sgRNAs would have different off-target effects from Cas9 nucleases or BE3, so the method of proving the accuracy only for ABEs was needed. To do this, Digenome-seq method, which identifies and captures DNA DSBs that are caused by Cas9 nucleases, was modified (Kim et al., 2015). One strand was cleaved by Cas9 nickases of ABEs and the other strand was cut using Endo V or hAAG/Endo VIII combination, which can recognize and inosine, the deamination product of A. After digesting the DNA of both strands, the whole genome was sequenced and aligned to the reference genome, and the Digenome cleavage score was calculated to obtain a potential off-target list.

From Digenome-seq results of 7 target sites, I could confirm that ABEs (8 ± 4 per target site, on average) were more specific than Cas9 (70 ± 30 sites (Kim et al., 2016)) and showed a potential off-target different from that of CBEs, And, unlike Cas9, DNA and RNA bulge off-targets were allowed as CBEs (Kim et al., 2017a). Also, I could validate that Digenome-seq could detect off-target sites with a substitution frequency of 0.1% or more in human cells.

To increase the specificity, I tested three methods. The first method was to modify 5 'of sgRNA (Cho et al., 2014; Fu et al., 2014; Kim et al., 2015; Kim et al.,

2016; Tsai et al., 2015). The second method was to use Sniper-Cas9, which was evolved to be specific instead of wild type Cas9 (Lee et al., 2018b). The last method was to use RNP delivery method instead of plasmid DNA delivery method (Kim et al., 2017b; Kim et al., 2014; Rees et al., 2017). Guide RNA modification is the truncation at 20-nt spacer length or extension of the spacer length of sgRNA with additional G or GG. In the case of truncation, the on-target activity was decreased as well and higher activity was observed in some off-targets. However, in the case of extended sgRNAs (gX_{20} or ggX_{20}), it was confirmed that the off-target effects were decreased while the on-target activity remained unchanged, so the specificity was increased.

Sniper-Cas9 is Cas9 which improves specificity like existing enhanced specificity Cas9 (eSpCas9), Cas9-high fidelity (Cas9-HF) or Hypa Cas9 specificity (Chen et al., 2017; Kleinstiver et al., 2016; Slaymaker et al., 2016) but it is made by selection system of random library unlike existing other specific Cas9 variants made by rational design. Unlike other specific Cas9 variants, Sniper-Cas9 has the advantage of working well with modified sgRNAs, so I used it for ABEs instead of wild type Cas9 and found that the specificity can be increased up to 5 times or more. If working with modified sgRNAs, it is possible to increase the specificity more than 80-fold.

In the case of RNP delivery methods, no external DNA is needed so GMO issues can be avoided in some cases (Kanchiswamy et al., 2015). In addition, since it has been reported that the off-target effects are low due to the short duration of expression due to the DNA-free delivery (Kim et al., 2014; Rees et al., 2017), the

RNP delivery method was used to increase the specificity and could be increased to about 7 times.

In summary, I showed, codon optimization and the number of NLSs is important for ABEs activity. Also, using mismatched sgRNAs, Digenome-seq capture, and validation off-targets in human cells, it was confirmed that ABE7.10, BE3, and Cas9 can recognize different sets of off-target sites in the human genome, calling for independent assessments of each tool. The seven ABE7.10 deaminases were reasonably specific, catalyzing A-to-G conversions *in vitro* at a limited number of sites in the human genome. I also showed that use of modified sgRNAs, RNPs, and Sniper ABE7.10 can reduce or avoid ABE off-target activity.

Reference

Barrangou, R. (2014). RNA events. Cas9 targeting and the CRISPR revolution. Science *344*, 707-708.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A *95*, 10570-10575.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V.W., Holt, L.*, et al.* (2016). GENOME ENGINEERING. The Genome Project-Write. Science *353*, 126-127.

Cameron, P., Fuller, C.K., Donohoue, P.D., Jones, B.N., Thompson, M.S., Carter, M.M., Gradia, S., Vidal, B., Garner, E., Slorach, E.M.*, et al.* (2017). Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nat Methods *14*, 600-606.

Cantor, C.R. (1990). Orchestrating the Human Genome Project. Science *248*, 49-51.

Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol Cell *47*, 497-510.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E, P.R.I., Lin, S., Kiani,

S., Guzman, C.D., Wiegand, D.J., *et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. Nat Methods *12*, 326-328.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature *550*, 407-410.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol *31*, 230-232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res *24*, 132-141.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., *et al.* (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods *10*, 361-365.

Cullot, G., Boutin, J., Toutain, J., Prat, F., Pennamen, P., Rooryck, C., Teichmann, M., Rousseau, E., Lamrissi-Garcia, I., Guyonnet-Duperat, V., *et al.* (2019). CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. Nat Commun *10*, 1136.

Eisenberg, D., Marcotte, E.M., Xenarios, I., and Yeates, T.O. (2000). Protein function in the post-genomic era. Nature *405*, 823-826.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol *33*, 179-186.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol *32*, 279–284

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature *551*, 464-471.

Gehrke, J.M., Cervantes, O., Clement, M.K., Wu, Y., Zeng, J., Bauer, D.E., Pinello, L., and Joung, J.K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat Biotechnol *36*, 977-982.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C.*, et al.* (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661.

Gopalappa, R., Suresh, B., Ramakrishna, S., and Kim, H.H. (2018). Paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption. Nucleic Acids Res *46*, e71.

Hess, G.T., Fresard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted

somatic hypermutation in mammalian cells. Nat Methods *13*, 1036-1042.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol *33*, 510-517.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol *31*, 227-229.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

Jiang, W., Feng, S., Huang, S., Yu, W., Li, G., Yang, G., Liu, Y., Zhang, Y., Zhang, L., Hou, Y.*, et al.* (2018). BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. Cell Res *28*, 855-861.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Kanchiswamy, C.N., Malnoy, M., Velasco, R., Kim, J.S., and Viola, R. (2015). Non-GMO genetically edited crop plants. Trends Biotechnol *33*, 489-491.

Kang, B.C., Yun, J.Y., Kim, S.T., Shin, Y., Ryu, J., Choi, M., Woo, J.W., and Kim, J.S. (2018). Precision genome engineering through adenine base editing in plants. Nat Plants *4*, 427-431.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I., and Kim, J.S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods *12*, 237-243, 231 p following 243.

Kim, D., Kim, D.E., Lee, G., Cho, S.I., and Kim, J.S. (2019). Genome-wide target specificity of CRISPR RNA-guided adenine base editors. Nat Biotechnol *37*, 430-435.

Kim, D., Kim, S., Kim, S., Park, J., and Kim, J.S. (2016). Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. Genome Res *26*, 406-415.

Kim, D., Lim, K., Kim, S.T., Yoon, S.H., Kim, K., Ryu, S.M., and Kim, J.S. (2017a). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat Biotechnol *35*, 475-480.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet *15*, 321-334.

Kim, K., Ryu, S.M., Kim, S.T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., and Kim, J.S. (2017b). Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol *35*, 435-437.

Kim, S., Bae, T., Hwang, J., and Kim, J.S. (2017c). Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. Genome Biol *18*, 218.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res *24*, 1012-1019.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490-495.

Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., and Liu, D.R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nature Biotechnology *36*, 843.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420-424.

Koo, B., Kim, D.-e., Kweon, J., Jin, C.E., Kim, S.-H., Kim, Y., and Shin, Y. (2018). CRISPR/dCas9-mediated biosensor for detection of tick-borne diseases. Sensors and Actuators B: Chemical *273*, 316-321.

Kwon, D.Y., Zhao, Y.T., Lamonica, J.M., and Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. Nat Commun *8*, 15315.

Lee, H., Choi, J., Jeong, E., Baek, S., Kim, H.C., Chae, J.H., Koh, Y., Seo, S.W.,

Kim, J.S., and Kim, S.J. (2018a). dCas9-mediated Nanoelectrokinetic Direct Detection of Target Gene for Liquid Biopsy. Nano Lett *18*, 7642-7650.

Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.H., Lee, K., Jung, I., Kim, D., Kim, S.*, et al.* (2018b). Directed evolution of CRISPR-Cas9 to increase its specificity. Nat Commun *9*, 3048.

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., Zhang, R., and Gao, C. (2018a). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. Genome Biol *19*, 59.

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X.*, et al.* (2018b). Base editing with a Cpf1-cytidine deaminase fusion. Nat Biotechnol *36*, 324-327.

Liang, F., Han, M., Romanienko, P.J., and Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A *95*, 5172-5177.

Liu, Z., Chen, M., Chen, S., Deng, J., Song, Y., Lai, L., and Li, Z. (2018a). Highly efficient RNA-guided base editing in rabbit. Nat Commun *9*, 2717.

Liu, Z., Lu, Z., Yang, G., Huang, S., Li, G., Feng, S., Liu, Y., Li, J., Yu, W., Zhang, Y.*, et al.* (2018b). Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. Nat Commun *9*, 2338.

Losey, H.C., Ruthenburg, A.J., and Verdine, G.L. (2006). Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. Nat Struct Mol Biol *13*, 153-159.

Ma, H., Tu, L.C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson,

T. (2016a). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat Biotechnol *34*, 528-530.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016b). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods *13*, 1029-1035.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct *1*, 7.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Mao, G., Marotta, F., Yu, J., Zhou, L., Yu, Y., Wang, L., and Chui, D. (2008). DNA context and promoter activity affect gene expression in lentiviral vectors. Acta Biomed *79*, 192-196.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y*., et al.* (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N.,

Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y.*, et al.* (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Rees, H.A., Komor, A.C., Yeh, W.H., Caetano-Lopes, J., Warman, M., Edge, A.S.B., and Liu, D.R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun *8*, 15790.

Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nature Reviews Genetics.

Rodriguez-Rodriguez, D.R., Ramirez-Solis, R., Garza-Elizondo, M.A., Garza-Rodriguez, M.L., and Barrera-Saldana, H.A. (2019). Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review). Int J Mol Med *43*, 1559-1574.

Ryu, S.M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.T., Kim, H.S., Kim, D.E., Lee, H., Chung, E.*, et al.* (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat Biotechnol *36*, 536-539.

Satomura, A., Nishioka, R., Mori, H., Sato, K., Kuroda, K., and Ueda, M. (2017). Precise genome-wide base editing by the CRISPR Nickase system in yeast. Sci Rep *7*, 2095.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc *3*, 1101-1108.

Shao, M., Xu, T.R., and Chen, C.S. (2016). The big bang of genome editing technology: development and application of the CRISPR/Cas9 system in disease animal models. Dongwuxue Yanjiu *37*, 191-204.

Shao, Y., Wang, L., Guo, N., Wang, S., Yang, L., Li, Y., Wang, M., Yin, S., Han, H., Zeng, L.*, et al.* (2018). Cas9-nickase-mediated genome editing corrects hereditary tyrosinemia in rats. J Biol Chem *293*, 6883-6892.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84-88.

Staahl, B.T., Benekareddy, M., Coulon-Bainier, C., Banfal, A.A., Floor, S.N., Sabo, J.K., Urnes, C., Munares, G.A., Ghosh, A., and Doudna, J.A. (2017). Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. Nat Biotechnol *35*, 431-434.

Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. Nat Methods *14*, 607-614.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P.*, et al.* (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.

Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D., and van Teeffelen, S. (2018).

Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. Mol Syst Biol *14*, e7899.

Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipniss, N.H., La Russa, M., and Qi, L.S. (2018a). CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. Cell *175*, 1405-1417 e1414.

Wang, X., Li, J., Wang, Y., Yang, B., Wei, J., Wu, J., Wang, R., Huang, X., Chen, J., and Yang, L. (2018b). Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. Nat Biotechnol *36*, 946-949.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J., and Yee, J.K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat Biotechnol *33*, 175-178.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov *2*, 16009.

Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L.*, et al.* (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat Commun *8*, 15058.

Yeo, N.C., Chavez, A., Lance-Byrne, A., Chan, Y., Menn, D., Milanova, D., Kuo, C.C., Guo, X., Sharma, S., Tung, A.*, et al.* (2018). An enhanced CRISPR repressor for targeted mammalian gene regulation. Nat Methods *15*, 611-616.

Zafra, M.P., Schatoff, E.M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A.Y.,

Simon, A., Han, T., Goswami, S., Montgomery, E.*, et al.* (2018). Optimized base editors enable efficient editing in cells, organoids and mice. Nat Biotechnol.

Zhang, Y., Massel, K., Godwin, I.D., and Gao, C. (2018). Applications and potential of genome editing in crop improvement. Genome Biol *19*, 210.

국문초록

CRISPR/Cas9 시스템은 최근 10 년간 생명 공학 분야에서 가장 널리 쓰이는 기술 중 하나이다. CRISPR 시스템은 박테리아 면역 체계에서 유래한 것으로 현재는 프로그래밍 가능한 핵산 분해 효소, 즉 유전체 교정 도구로서 이용된다. 이 유전체 교정 도구는 DNA 이중 가닥 절단을 만든 후 세포 복구 기작을 이용하여 유전자를 망가뜨릴 수 있을 뿐 아니라 외부 유래 DNA와 함께 작용하여 유전자를 삽입하는 데에도 사용될 수 있다. 기존의 CRISPR 시스템은 정교한 상동재조합 방법보다는 비상동성 말단 접합 방법으로 복구가 일어나므로 주로 유전자를 망가뜨리는 데 사용된다는 한계가 있다.

그러나 인간의 질병을 일으킬 수 있다고 알려진 돌연변이의 약 58%는 점 돌연변이다. 따라서 점 돌연변이를 도입하는 기술의 필요성이 제기되었고, 최근 이를 가능하게 하는 CRISPR 염기교정 유전자 가위가 개발되었다. CRISPR 염기교정 유전자 가위는 DNA 이중 가닥을 자르는 아미노산이 일부 비활성화된 Cas9 (Cas9 니케이즈)과 시토신/아데닌 탈아미노 효소로 이루어져 있다. 시토신 염기교정 유전자 가위는 C:G 염기쌍을 T:A 염기쌍으로 바꿀 수 있으며, 아데닌 염기교정 유전자 가위는 아데노신을 탈아미노화 하여 A:T 염기쌍을 G:C 염기쌍으로 전환시킬 수 있다. 이러한 염기교정 유전자 가위는 최근 여러 종류가 독립적으로 보고되었으며 박테리아, 인간 세포, 동물 및 식물 등에 적용되고

있지만 염기 교정 유전자 가위, 특히 아데닌 염기 교정 유전자 가위의 특이성에 대한 연구는 기존에 없었다.

이 논문에서는 우선 아데닌 염기 교정 유전자 가위의 염기 교정 효율이 코돈 사용에 따른 발현량에 영향을 받고 핵 위치 신호의 수가 많아진다면 효율이 높아질 수 있다는 것을 입증하였다. 또한 DNA 이중 가닥 절단 위치를 인식하는 방법으로 알려진 기존의 Digenome-seq 방법을 변형하여 아데닌 염기 교정 유전자 가위의 비표적 효과를 검증하는 방법을 소개하였다. 아데닌 염기 교정 유전자 가위의 비표적 효과는 Cas9 DNA 이중 가닥 절단효소나 시토신 염기 교정 유전자가위의 비표적 효과와는 다르다는 것을 확인할 수 있다. 이러한 비표적 효과가 인간 세포에서도 실제로 발생하는지를 확인하고 마지막으로는 DNA와 염기쌍을 이룰 수 있는 spacer 염기 서열의 길이를 변형시킨 sgRNAs를 사용하거나 와일드 타입의 Cas9 니케이즈 대신 특이적이라고 알려진 Sniper Cas9 니케이즈를 사용하는 방법 외에도 플라스미드 DNA 전달 대신 단백질과 전사된 RNA 형태로 전달하는 방법이 아데닌 염기 교정 유전자 가위의 특이성을 증가시키는데 효과적이라는 것을 보였다.

학 번: 2014-21245